

## 한국자료분석학회 2023년도 하계 학술논문발표대회 준비위원

### 조직위원회

위원장	한상태	호서대학교 빅데이터AI학부
부위원장	이학배	연세대학교 응용통계학과
	김기환	고려대학교 빅데이터사이언스학부
	김용대	서울대학교 통계학과
	홍재범	부경대학교 경영학부
	이영섭	동국대학교 통계학과
	윤성민	부산대학교 경제학부
	박찬근	한국해양대학교 데이터사이언스전공
	박승열	(주)케이스탯리서치 회장
위원	이성건	성신여자대학교 수리통계데이터사이언스학부
	조형준	고려대학교 통계학과
	최호식	서울시립대학교 도시빅데이터융합학과
	진서훈	고려대학교 빅데이터사이언스학부
	허태영	충북대학교 정보통계학과
	이동희	경기대학교 경영학부
편집위원장	강현철	호서대학교 빅데이터AI학부
자문위원	최종후	고려대학교 응용통계학과
	조완현	전남대학교 통계학과
	박희창	창원대학교 통계학과
	강창완	동의대학교 산업경영·빅데이터공학과

### 운영위원회

공동위원장	조형준	고려대학교 통계학과
	최호식	서울시립대학교 도시빅데이터융합학과
부위원장	이성건	성신여자대학교 수리통계데이터사이언스학부
	신승준	고려대학교 통계학과
	김지수	가천대학교 간호학과
	제상영	고려대학교 경제통계학과
	조장식	경성대학교 정보통계학과
위원	김지희	강원대학교 응급구조학과
	김태훈	경성대학교 경제금융물류학부
	김성환	건국대학교 응용통계학과
	김양진	숙명여자대학교 통계학과
	곽일엽	중앙대학교 통계학과
	박만식	성신여자대학교 수리통계데이터사이언스학부
	김은석	(주)GDS 컨설팅그룹 대표이사
	김지연	(주)케이스탯리서치 대표이사



# 한국자료분석학회

## 2023년도 하계 학술논문발표대회

- 일시 : 2023년 7월 6일(목) ~ 7월 7일(금)
- 장소 : 고려대학교(서울특별시 성북구 안암로 145)
- 주최/주관 : 한국자료분석학회, Data Science PLUS, 고려대학교 BK21통계학교육연구팀
- 후원 : 한국연구재단, (주)케이스태리서치, (주)GDS컨설팅, 나이스평가정보, 베가스, SAS 코리아, 데이터솔루션, 정보통신정책연구원

### ● 7월 6일 목요일

17:00 - 19:00 편집위원회(수당패컬티하우스 2층)

17:00 - 19:00 이사회(수당패컬티하우스 2층)

### ● 7월 7일 금요일

09:00 - 18:00 포스터발표(정경관 5층)

09:30 - 11:00 논문발표 A1 - A4(정경관 5층 501, 503, 507, 508호)

11:10 - 12:10 튜토리얼(정경관 5층 506호)

좌장 : 조형준교수(학술이사, 고려대)

제목 : 방송사 출구조사 방법론 - 20대 대선 출구조사를 중심으로

강사 : 김영원교수(숙명여자대학교 통계학과)

12:10 - 12:30 총회 (정경관 5층 506호)

진행 : 이성건교수(총무이사, 성신여대)

12:30 - 13:40 중식시간(수당패컬티하우스 3층)

13:40 - 14:40 학회장 초청 강연(정경관 5층 506호)

좌장 : 한상태교수(학회장, 호서대)

제목 : 삶의 변화 : 돌아보다, 지켜보다, 내다보다

강사 : 송길영부사장(바이브컴퍼니, 구 다음소프트)

14:50 - 16:20 논문발표 B1 - B5(정경관 5층 501, 503, 506, 507, 508호)

16:30 - 18:00 논문발표 C1 - C5(정경관 5층 501, 503, 506, 507, 508호)



**논문발표 A1 (정경관 501호)**

Data Science PLUS 기획세션

좌장 : 양대원(충남대학교)

---

09:30	1. 모바일 헬스케어 상담알고리즘 개발 .....	1
	한상태*, 강현철, 연규필(호서대), 최호식(서울시립대)	
11:00	2. Distributional Learning of Variational AutoEncoder: Application to Synthetic Data Generation .....	3
	안승환*, 전종준(서울시립대학교 통계학과)	
	3. Bibimbap : Ensembling Diverse Pre-trained Models for Domain Generalization in Domain Shifted Task .....	5
	Jinho Kang*(Department of Artificial Intelligence, University of Seoul), Taero Kim(Department of Applied Statistics, Department of Statistics and Data Science, Yonsei University), Jiyoung Jung(Department of Artificial Intelligence, University of Seoul), Rakwoo Chang(Department of Chemistry, University of Seoul), Kyungwoo Song(Department of Artificial Intelligence, University of Seoul)	
	4. Nonparametric Bayesian Sparse Latent Factor Model for Multivariate Functional Clustering: an Application to Multiple Air Pollutants Data .....	7
	양대원*(충남대학교), 최태린(고려대학교), Eric Lavigne(School of Epidemiology and Public Health, University of Ottawa), 정연승(한국과학기술원)	

**논문발표 A2 (정경관 503호)**

고려대학교 BK21  
통계학교육연구팀 기획세션

좌장: 김경희(고려대학교)

---

09:30	1. Rank-based Regression for Doubly-Interval-Censored Data .....	9
	박서현, 최상범(고려대학교 통계학과), Wenbin Lu (Department of Statistics, NC State University), Zhezhen Jin (Department of Biostatistics, Columbia University)	
11:00	2. A Unified Principal Sufficient Dimension Reduction R Package : psdr .....	11
	신정민*, 신승준(고려대학교 통계학과)	
	3. Unpaired Regression for a Discrete Response via Poisson Quantiles Matching .....	13
	임형준*, 김경희(고려대학교 통계학과)	
	4. 절단자료를 위한 의사결정나무 .....	15
	김정환*, 조형준(고려대학교 통계학과)	

**논문발표 A3 (정경관 507호)**

**의료/간호/보건**

좌장 : 유제복(경상국립대학교)

- 
- 09:30 | 1. Pup Mad 논문 정도 인지 장애 중국 침 치료 중재 효과 코크란 메타분석 .....  
11:00 | 김원순\*(한국융합간호학 연구소)
2. 70개 중진료권에 대한 응급실 방문 환자의 유형과 연관성 ..... 17  
허영진\*, 김세형, 이한나, 오미라(국립중앙의료원)
3. Ensemble Learning for COVID-19 Detection on CT Scan Images ..... 19  
Van-Phong Truong\*(Department of Data Informatics, National Korea Maritime and Ocean University), Thi-Ngot Pham(Interdisciplinary Major of Ocean Renewable Energy Engineering, National Korea Maritime and Ocean University), Jun-Ho Huh, Chan-Keun Park(Department of Data Sciences, National Korea Maritime and Ocean University)
4. COVID-19 팬데믹 상황에서 간호대학생의 온라인 취업적응프로그램의 효과 ..... 25  
유제복\*(경상국립대학교 간호대학)

**논문발표 A4 (정경관 508호)**

**경제경영**

좌장 : 김인중(홍익대학교)

- 
- 09:30 | 1. CEO Pay-performance Sensitivities in the ESG Managed Firms ..... 29  
11:00 | 김인중\*(홍익대학교 금융보험학과)
2. 호텔·외식 고객의 기술불안과 행동의도 간의 영향관계: 셀프서비스기술 서비스품질과 고객만족의 다중매개효과 ..... 33  
문혜영\*(연세대학교 심바이오틱 라이프텍 연구소)
3. 투자지표로서 거래량의 유의성에 관한 연구 ..... 39  
김영준\*, 임상빈, 양기성(숭실대학교)

**논문발표 B1 (정경관 501호)**

Data Science PLUS 기획세션

좌장 : 김영선(고려대학교)

---

14:50	1. Multilevel Topic Model : a Nonparametric Multilevel Latent Class Analysis for Topic Dependency .....	41
16:20		
	김영선*, 정환(고려대학교), 전새봄(목원대학교)	
	2. Robust Support Vector Machine for Multivariate Functional Data .....	43
	심정은*, 신승준(고려대학교 통계학과)	
	3. On Weighted Least Squares Regression with Partially Interval-censored Data .....	45
	김예지*, 최상범(고려대학교 통계학과)	
	4. 모바일 게임에서 이탈 조짐 유저에 대한 조기 탐지 .....	47
	이민혁*, 박선우, 이성환, 김수인, 정운서(고려대학교 통계학과)	

**논문발표 B2 (정경관 503호)**

통계

좌장 : 곽일엽(중앙대학교)

---

14:50	1. Proformer : a Hybrid Macaron Transformer Model Predicts Expression Values from Promoter Sequences .....	49
16:20		
	곽일엽*, 김병찬, 이주현(중앙대), Daniel Garry, Wuming Gong(University of Minnesota)	
	2. 클래스 불균형 데이터 분류를 위한 군집화 기반 언더샘플링 연구 .....	51
	김주미*, 정여진(국민대학교)	
	3. Mixture Modelling for Analyzing a Rainfall Pattern with Julia: A Case Study in South Korea .....	53
	Hee-Young Kim(Bigdata Science, Korea University), Su-Hyeon Kim, Ji-Yeon Oh*(Department of National Statistics, Korea University)	
	4. Imputation Methods for Multivariate Longitudinal Data Using Mixture of Generalized Linear Mixed Models .....	55
	양대경*, 송주원(고려대학교 통계학과)	

**논문발표 B3 (정경관 506호)**

**정보보호**

좌장 : 최호식(서울시립대학교)

---

14:50	1. Introduction to Privacy Preserving Data Sharing and Analytics with Homomorphic Encryption .....	57
16:20		
	2. Privacy Preserving Machine Learning for Secure Outsourced Computing .....	59
		김설기*, 어수행, 이가람(크립토크랩)
	3. Application of Homomorphic Encryption in Predicting Postoperative Mortality using Multi-institutional Data .....	61
		이가람*(크립토크랩), 서준교(서울아산병원 비뇨의학과), 이상욱(서울아산병원 마취통증의학과), 신준범, 김설기(크립토크랩)
	4. Land Use Regression Method as Sign Constrained Regularized Regressions .....	63
		권순선*(아주대학교 수학과), 최호식(서울시립대 도시빅데이터융합학과), 이환희(부산대학교 의생명융합공학부), 김연진(서울대학교 보건대학원), 김환철(인하대학교 직업환경의학과), 이우주(서울대학교 보건대학원)

**논문발표 B4 (정경관 507호)**

**의료/간호/보건**

좌장 : 전은미(배재대학교)

---

14:50	1. 간호대학생의 비판적 사고성향과 간호정보역량이 간호과정 수행능력에 미치는 영향 .....	65
16:20		
		김민영*(경남정보대학교)
	2. 다문화청소년의 문화적응스트레스가 학교생활에 미치는 영향: 친구의 지지와 자아 존중감의 이중매개효과 .....	69
		박공주*(김해대학교)
	3. 골다공증 여성노인의 건강관련 삶의 질 영향요인: 국민건강영양조사 자료 활용 (2016~2020년) .....	73
		전은미*(배재대학교), 강창완(동의대학교)
	4. 병기에 따른 관측중단 손실의 영향 감소 조정 DSM을 이용한 NSCLC 생존분석 75	
		박소연*(전남대학교 수학과/통계학과), 김민수(전남대학교 통계학과), 강세령(화순전남대학교병원 핵의학과)



**논문발표 B5 (정경관 508호)**

교육/심리/사회/환경/교통

좌장 : 김태영(동의대학교)

---

14:50	1. 디퍼러닝 연계 교양교육튜터링 효과성 분석 .....	77
	김태영*, 윤혜경, 최승배(동의대학교)	
16:20	2. cs-QCA를 활용한 환경분쟁조정 인과성분석: 층간소음 사례연구 .....	79
	송경렬*(한국외국어대학교 국정관리연구소)	
	3. 교통 네트워크 데이터에 대한 거대언어모델 활용 방안 연구 .....	83
	정재경*, 김예은, 이다운, 최호식(서울시립대학교 도시빅데이터융합학과)	

**논문발표 C1 (정경관 501호)**

Data Science PLUS 기획세션 좌장 : 송준(고려대학교 통계학과)

---

16:30	1. Robust Graph Augmentation for Molecular Property Prediction .....	85
	조소영*, 홍성철(서울시립대 통계학과),	
18:00	임창원(중앙대 응용통계학과), 전종준(서울시립대 통계학과)	
	2. Sparse Sufficient Dimension Reduction via Outer Product of Gradients using Composite Quantile Regression .....	87
	이석영*, 신승준(고려대학교 통계학과)	
	3. 앙상블 모델을 이용한 MMORPG의 붓 탐지 연구 .....	89
	정현수*, 최정원, 원상준, 김수인, 정운서(고려대학교 통계학과)	
	4. Functional Adaptive Group Lasso with its Non-asymptotic Bounds .....	91
	장세훈*, 송준(고려대학교 통계학과)	

**논문발표 C2 (정경관 503호)**

통계

좌장 : 송광윤(조선대학교)

- 
- 16:30 | 1. 일반화된 지수 분포를 고려한 NHPP 소프트웨어 신뢰성 모형에 관한 연구 ..... 93  
 18:00 | 송광윤\*, 장인홍, 김윤수, 이다혜(조선대학교)
2. 확률과정기법을 이용한 사육 환경 변화에 따른 한우 증체량 예측 ..... 95  
 강소라\*(전남대, 수학/통계학과), 조완현, 나명환(전남대 통계학과)
3. 감성분석에서의 BERT와 GPT 모형 비교(D대학교 서술형 강의 평가 분석) ..... 99  
 성진용\*(동의대학교), 권민수(에스티이노베이션 대표이사), 심성현, 최승배,  
 강창완(동의대학교 산업경영빅데이터공학과)
4. 딥러닝 기반 시계열 분석 모형의 불확실성 정량화 비교 연구 ..... 101  
 윤영인\*(한양대학교 응용수학과), 정혜영(한양대학교 수리데이터사이언스학과)

**논문발표 C3 (정경관 506호)**

경제경영통계

좌장 : 전새봄(목원대학교)

- 
- 16:30 | 1. 중국의 호주산 석탄 수입 금지가 원료탄 가격에 미친 영향: 반사실적 분석을 중심  
 18:00 | 으로 ..... 103  
 김민수\*, 임석, 유지애(코리아피디에스)
2. 다중 스펙트럼 영상과 Video Vision Transformer를 이용한 버의 수확량 예측 .... 105  
 김다현\*(전남대학교 수학/통계학과), 나명환(전남대학교 통계학과)
3. 국민건강보험빅데이터를 이용한 한국의 세대 간 소득이동성 연구 ..... 107  
 전새봄\*(목원대학교 마케팅빅데이터학과), 김준일(목원대학교 금융경제학과)

**논문발표 C4 (정경관 507호)**

통계

좌장 : 김동하(성신여자대학교)

---

16:30	1. A Study on Efficient Data Synthesis Using Deep Generative Models .....	109
		김지우*, 박세리, 김동하(성신여자대학교)
18:00	2. YOLO와 XAI를 활용한 궤양병 분류 알고리즘 연구 .....	111
		장진결*, 김기환(고려대학교)
	3. Correlation-Based Clustering Analysis of the Opioid Epidemic Data on East Texas Counties .....	113
		Doo Young Kim*(Sam Houston State University)
	4. Improving the Usability of Collaborative Robot based on Data Analysis .....	115
		정소영*, Christoffer Sloth(University of Southern Denmark)

**논문발표 C5 (정경관 508호)**

기업세션

좌장 : 김정환(고려대학교)

---

16:30	1. 기업평가를 위한 기업정보의 다각적 활용 사례 .....	117
		이연경*, 김종윤(NICE평가정보)
18:00	2. 제품 불량 검출을 위한 딥러닝 모델 개발 사례 .....	119
		김민정*(테이타솔루션)
	3. 사회안전지수와 지역주민 인식에 대한 분석 .....	121
		박현수(충북대학교 국가위기관리연구소), 장안식*(케이스탯컨설팅 공공사회정책연구소)
	4. IPA 활용 지역수출유망단지 진단평가 .....	123
		홍우진*(전라남도농업기술원), 손황배(농촌진흥청)

1. 서울 아파트 가격에 영향을 주는 요인들 .....	127
심성훈*(부경대학교 경제학과)	
2. 토픽모델링을 활용한 기술동향 파악: AI 관련 특허를 중심으로 .....	133
김용우*, 김민구, 김정희(한양대학교 기술경영전문대학원 기술경영학과)	
3. 2단계 집락 조건부 이산양적모형 .....	137
이기성*(우석대학교 아동사회복지학부), 홍기학(동신대학교 컴퓨터학과), 손창균(동국대학교 빅데이터·응용통계학과 교수), 박근화(한국문화관광연구원), 홍성준(마사회 말산업연구소)	
4. 자료 포락 분석(DEA)을 이용한 스타트업 기술 효율성 예측(머신러닝) 모델: 민간주도형 기술창업지원 팀스(TIPS) 선정 기업을 중심으로 .....	141
박현민*, 김정호, 오주희(한동대학교)	
5. An Acoustic Feature Based Popularity Prediction Using Machine Learning Approach: Indie Music Scene .....	143
조은비*(한동대학교 경영경제학부), 장하은(한동대학교 커뮤니케이션학부), 오주희(한동대학교 경영경제학부)	
6. TVAE를 이용한 영 과잉 데이터 생성 연구 .....	145
길경주*(서울시립대학교)	
7. 역확률 가중치 추정량과 그래프 라쏘를 이용한 고차원 자료에서 다중대체를 위한 변수 선택방법 제안 .....	147
이윤아*, 박성오(성신여자대학교 통계학과)	
8. 소상공인 자립을 위한 서울시 행정동별 폐업 리스크 예측 모델 .....	149
송예진*(성신여자대학교 통계학과), 박지호, 백예은, 최정빈(성신여자대학교 수리통계데이터사이언스학부), 김동하(성신여자대학교 통계학과, 수리통계데이터사이언스학부)	
9. DBSCAN을 활용한 구성자료 군집분석 .....	151
이제석*, 김병원(경북대학교 통계학과)	
10. Box Monitoring of Mixed-type Multivariate Data using Gower-type Distance .....	153
김우현*, 이성임(단국대학교), 임요한(서울대학교)	
11. LDA를 이용한 HMD 기반 가상 현실 장치 사용자 경험 분석 연구 .....	155
유일선*(동덕여자대학교)	
12. 앙상블 비모수적 변수 선택 방법 : 위게임 데이터 분석 .....	157
신웅섭*, 한봉규, 차영호(육군 전투지휘훈련단)	
13. 중국 대학생의 연애관 분석 .....	159
평인란*, 양정호(성균관대 교육학과)	

14. Analysis of Chinese high school students' motivation to study abroad and it's external factors .....	161
Jin Hua*, Jung-Ho Yang(성균관대학교)	
15. 중국 대학의 국제화 지표 및 순위 분석 .....	163
왕호*, 양정호(성균관대학교)	
16. 중국 교원순환 기준 데이터 분석 .....	165
엽혜림*, 양정호(성균관대학교)	
17. 중국 가오카오이민 관련 입학자료 분석 .....	167
손옥하*, 양정호(성균관대학교)	
18. 원거리 시각 초점 고정과 어지러움증과의 관계 .....	169
노효련*, 전상준, 홍성현, 안소현, 구교준, 이유진, 이수빈(강원대학교 물리치료학과) 이영희(영산대학교 미용건강관리학과)	
19. 국내·외 악력 측정기의 악력 차이 분석 .....	173
노효련*, 김다정, 이한민, 이지은, 임철균, 박윤아, 김혜미(강원대학교 물리치료학과), 이영희(영산대학교 미용건강관리학과), 이효택(바이오헬스케어 대표)	
20. 멀미 증상과 사이버 멀미와의 상관관계 .....	177
노효련*, 김혜규, 남채림, 한재영, 이효정, 최중현, 김건우, 김유연(강원대학교 물리치료학과)	
21. 원예요법을 병합한 대인간호 참여 경험 .....	181
박송이*(강동대학교 간호학부), 차지영(이화여자대학교 간호대학)	
22. 간호학과 신입생의 대인관계역량이 대학생활적응에 미치는 영향: 사회적 관계망 접촉빈도의 매개효과를 중심으로 .....	183
이경은*(계명문화대학교 간호학과)	
23. Importance of Multimorbidity in Predicting Suicidal Ideation Among the Aging Population Using Machine Learning Methods .....	189
서영빈*(고려대학교 헬스케어사이언스학과), 김혜영(고려대학교 보건과학과), 최기봉(고려대학교 보건정책관리학부)	
24. 간호대학생의 감성지능, 의사소통능력, 대인관계유능성 및 대학생활적응의 관계 .....	191
이은미*, 오윤정(남부대학교 간호학과)	
25. 간호대학생의 인권의식, 공감능력 및 다문화수용성과의 관계 .....	193
오윤정*, 이은미(남부대학교 간호학과)	



## 튜토리얼

(1) 발표자 : 김영원(숙명여대 통계학과)

(2) 제목 : 방송사 출구조사 방법론 - 20대 대선 출구조사를 중심으로

오늘날 방송사 출구조사의 성공은 하루아침에 이루어진 것이 아니다. 선거예측조사 도입 초창기에는 참혹한 오류를 저지르기도 했으나, 통계학자들이 본격적으로 조사설계에 참여한 2006년 지방선거 출구조사 이후 현재의 과학적인 출구조사 방법론이 자리 잡게 되었고, 이제는 출구조사 자체를 선거제도의 일부인 것처럼 유권자들은 인식하게 되었다.

이번 강연에서는 21대 대통령선거 출구조사 방법론을 중심으로 출구조사 과정에서 발생할 수 있는 다양한 형태의 표집오차와 비표집오차 발생 요인과 이런 오차를 최소화하기 위한 표본 투표소 추출 방법, 표본 투표소 내에서 조사 대상 투표자 선정 및 실사 과정, 응답 거절자 처리 방법, 15시 이후 투표자 보정, 국회의원선거 정당별 의석수 예측 문제 등 현재 지상파 방송 3사 출구조사에 적용되고 있는 조사설계와 분석과정을 구체적으로 살펴보기로 한다. 특히 2014년 지방선거부터 본격 시행된 ‘사전투표제’가 보편화되면서 급격히 늘어나는 사전투표자 분석을 위해 20대 대통령선거 출구조사에 새로 도입한 전화조사를 토대로 한 사전투표자 득표율 예측과정도 설명한다.

## 초청강연

(1) 발표자 : 송길영(바이브컴퍼니 부사장, 구 다음소프트)

(2) 제목 : 삶의 변화 : 돌아보다, 지켜보다, 내다보다

기술의 발달로 이룬 초연결과 자동화는 우리의 삶을 변화시킨다. 살아가는 것이 살아보는 것으로 변화하는 것은 질곡의 삶이 취향의 삶으로 업그레이드 되는 시점에서 누릴 수 있는 특권과 같다. 그 결과로 인류는 물질의 풍요로움을 누릴 수 있게 해 주었으나 그 이상의 정서적 풍요로움을 추구하고자 하는 사람들의 변화는 더 많은 혁신과 배려를 요구하고 있다. 해야할 일을 대행해주고 하고 싶은 일을 추구할 수 있도록 돕는 삶의 변화 속, 새로운 기회를 어떻게 함께 이룰 것인가 고민해 본다.



## 모바일 헬스케어 상담알고리즘 개발

한상태<sup>1</sup>, 강현철<sup>2</sup>, 연구필<sup>3</sup>, 최호식<sup>4</sup>

### 요 약

보건소 모바일 헬스케어 사업은 기존의 시·공간적 한계를 극복함으로써 보건소 방문을 최소화 하고, 모바일 앱(APP)과 디바이스를 통해 개인의 건강생활 실천정보를 실시간으로 수집하여 맞춤형 건강관리 서비스를 제공하고자 함이 목적이다. 16년도에 1차 시범사업으로 10개 보건소를 통해 그 효과를 입증하였고, 이에 따라, 1차년도 시범사업 모형이 전국 확산 시에도 적용 가능한 지 검증하고자 35개 보건소로 확대하여 운영하고, 향후 전국으로 확산할 수 있는 기반을 마련하고자 하였다. 보건소 전문 인력들이 대상자의 활동 정보를 영역별(건강, 신체활동, 영양)로 일일이 분석하여 매달 맞춤형 상담 코멘트를 제공해야 하므로 업무 과중 및 인력 부족 현상이 심각한 상황이다. 또한, 보건소별 전문 인력의 성향이나 역량에 따라 대상자 맞춤 상담 코멘트의 품질 수준 차이가 크게 발생하여 우수 상담 코멘트를 중심으로 코멘트의 품질을 향상하는 것이 매우 중요하게 대두되었다. 이러한 상황에서 2016년부터 축적된 건강정보 데이터와 양질의 상담코멘트 내용에 대한 분석을 통해 영역별 표준화된 상담알고리즘을 개발하여 모바일 헬스케어 사업의 질적 고도화를 이루고자 한다.

주요용어 : 모바일 헬스케어, 영역별 정보, 상담코멘트 내용, 상담알고리즘 개발.

<sup>1</sup>31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부 교수.

<sup>2</sup>31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부 교수.

<sup>3</sup>31499 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부 부교수.

<sup>4</sup>02504 서울시 동대문구 서울시립대로 163, 서울시립대학교 일반대학원 도시빅데이터융합학과 부교수.

## Distributional Learning of Variational AutoEncoder: Application to Synthetic Data Generation

*Seunghwan An*<sup>1</sup>, *Jong-June Jeon*<sup>2</sup>

### Abstract

The Gaussianity assumption has been consistently criticized as a main limitation of the Variational Autoencoder (VAE), despite its efficiency in computational modeling. In this paper, we propose a new approach that expands the model capacity (i.e., expressive power of distributional family) without sacrificing the computational advantages of the VAE framework. Our VAE model's decoder is composed of an infinite mixture of asymmetric Laplacian distribution, which possesses general distribution fitting capabilities for continuous variables. Our model is represented by a special form of a nonparametric M-estimator for estimating general quantile functions, and we theoretically establish the relevance between the proposed model and quantile estimation. We apply the proposed model to synthetic data generation, and particularly, our model demonstrates superiority in easily adjusting the level of data privacy.

*Keywords* : Variational AutoEncoder, distributional learning, synthetic data generation.

---

<sup>1</sup>Integrated PhD program, Department of Statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul, 02504, Korea. E-mail: dkstmdghks79@uos.ac.kr

<sup>2</sup>(Corresponding Author) Professor, Department of Statistics, University of Seoul, 163 Seoulsiripdaero, Dongdaemun-gu, Seoul, 02504, Korea. E-mail: jj.jeon@uos.ac.kr

## Bibimbap : Ensembling Diverse Pre-trained Models for Domain Generalization in Domain Shifted Task

강진호<sup>1</sup>, 김태로<sup>2</sup>, 김예원<sup>3</sup>, 오창대<sup>4</sup>, 정지영<sup>5</sup>, 장락우<sup>6</sup>, 송경우<sup>7</sup>

### 요약

이 연구는 훈련 데이터와 실제 데이터의 분포 차이에서 발생하는 머신 러닝의 문제를 해결하고자 한다. 딥러닝 모델의 강건성을 강화하고 성능을 개선하기 위해, 우리는 사전 학습 모델을 사용하여 앙상블 하는 앙상블 학습 프레임워크 'Bibimbap'을 제안한다. 우리는 기존의 가중치 앙상블 프레임워크에서 앙상블되는 모델이 각각 다른 손실 분지에 위치할 수 있음을 실험적으로 확인하였고, 이는 앙상블 후에 손실 장벽에 모델이 위치할 수 있음을 의미한다. 우리는 가중합 앙상블을 진행한 후 한번 더 모델 미세 조정 단계를 배치함으로써 손실 장벽으로 인해 발생할 수 있는 과소적합 문제를 해결하고 배치 정규화의 실행 매개변수를 안정화하였다. 우리는 또한 정성 분석단계를 통해 다양한 앙상블 모델을 사용하는 것이 도메인 일반화에 긍정적인 영향을 끼친다는 것을 확인하였다. 우리의 실험은 훈련-검증 데이터의 분할 방식으로 인해 극단적인 도메인 변화를 위한 실험에 적합한 화학 분자 데이터에서 진행하였다.

주요용어 : 전이 학습, 도메인 일반화, 가중치 평균, 앙상블 학습, 분자 분류.

\*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NO.2022M3J6A1084845, 2021R1F1A1060117, 2020R1C1C1008726)

<sup>1</sup>02054 대한민국 서울특별시 서울시립대로 163 서울시립대학교 인공지능학과 석사과정.  
E-mail: bubble3jh@uos.ac.kr

<sup>2</sup>03722 대한민국 서울특별시 서대문구 연세로 50 연세대학교 응용통계학과 석사과정.  
E-mail: taero.kim@yonsei.ac.kr

<sup>3</sup>02054 대한민국 서울특별시 서울시립대로 163 서울시립대학교 인공지능학과 석사과정.  
E-mail: yeyewon12@uos.ac.kr

<sup>4</sup>02054 대한민국 서울특별시 서울시립대로 163 서울시립대학교 인공지능학과 석사과정.  
E-mail: changdae.oh@uos.ac.kr

<sup>5</sup>02054 대한민국 서울특별시 서울시립대로 163 서울시립대학교 인공지능학과 교수.  
E-mail: jyjung@uos.ac.kr

<sup>6</sup>(교신저자)02054 대한민국 서울특별시 서울시립대로 163 서울시립대학교 융합응용화학학과 교수.  
E-mail: rchang90@uos.ac.kr

<sup>7</sup>(교신저자)03722 대한민국 서울특별시 서대문구 연세로 50 연세대학교 응용통계학과 교수.  
E-mail: kyungwoo.song@yonsei.ac.kr

## Nonparametric Bayesian covariate-dependent multivariate functional clustering: an application to multiple air pollutants data \*

*Daewon Yang*<sup>1</sup>, *Taeryon Choi*<sup>2</sup>, *Eric Lavigne*<sup>3</sup>, *Yeonseung Chung*<sup>4</sup>

### Abstract

Air pollution is a major threat to public health. Understanding the spatial distribution of air pollution concentration is of great interest to government or local authorities, as it informs about target areas for implementing policies for air quality management. Cluster analysis has been popularly used to identify groups of locations with similar profiles of average levels of multiple air pollutants, efficiently summarising the spatial pattern. This study aimed to cluster locations based on the seasonal patterns of multiple air pollutants incorporating the location-specific characteristics such as socio-economic indicators. For this purpose, we proposed a novel non-parametric Bayesian sparse latent factor model for covariate-dependent multivariate functional clustering. Furthermore, we extend this model to conduct clustering with temporal dependency. The proposed methods are illustrated through a simulation study and applied to time-series data for daily mean concentrations of ozone ( $O_3$ ), nitrogen dioxide ( $NO_2$ ), and fine particulate matter ( $PM_{2.5}$ ) collected for 25 cities in Canada in 1986-2015.

*Keywords* : Bayesian Nonparametrics; Covariate-dependent clustering; Dirichlet process; Indian Buffet process; Multivariate functional data; temporal dependency.

---

\*This research was supported by the Senior Research grant (2019R1A2C1086194) from the National Research Foundation (NRF) of Korea funded by the Ministry of Science, Information and Communication Technologies, the Basic Science Research grant (2019R1A2C1010018) from the NRF of Korea funded by the Ministry of Education, and the Government-wide R & D Fund project for Infectious Disease (HG18C0025).

<sup>1</sup>34134 Department of Information Statistics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon. Department of Information and Statistics, Chungnam National University.

<sup>2</sup>Department of Statistics, Korea University, Seoul, South Korea.

<sup>3</sup>School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada.

<sup>4</sup>Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon, South Korea.

## Rank-based Regression for Doubly-Interval-Censored Data<sup>\*</sup>

*Seohyeon Park*<sup>1</sup>, *Sangbum Choi*<sup>2</sup>, *Wenbin Lu*<sup>3</sup>, *Zhezhen Jin*<sup>4</sup>

### Abstract

In many biomedical fields, especially in studies of disease progressions, we frequently encounter two sequential events where both event times tend to be interval-censored due to regular examinations. Such a structure is called doubly interval censoring (DIC), and our primary interest is the elapsed time between two consecutive events. In this research, we propose a weighted rank regression approach for DIC data under the semiparametric accelerated failure time model. After transforming DIC data into simple interval-censored data where true elapsed times may lie, we develop estimating procedures with a Gehan-type weight by gathering all comparable pairs of observed residuals from transformed data. Moreover, we generalize it with data-dependent weights and extend it to clustered DIC data where the cluster size is potentially informative. An efficient resampling technique for the variance estimation is considered. We establish asymptotic properties and conduct numerical studies to demonstrate finite sample performances. Finally, we illustrate our method with dental data from Signal Tandmobiel study® to examine the effect of covariates on time to caries of four permanent first molars.

*Keywords* : Accelerated lifetime, Clustered failure time, Doubly interval-censoring, Gehan statistics, Survival analysis.

---

<sup>\*</sup>This work was in part supported by the National Research Foundation (NRF) of Korea, under grant 2022M3J6A1063595, 2022R1A2C1008514.

<sup>1</sup>Graduate Student, Department of Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: seohyeon20@korea.ac.kr

<sup>2</sup>(Corresponding Author) Associate Professor, Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: choisang@korea.ac.kr

<sup>3</sup>Professor, Department of Statistics, NC State University, Raleigh, North Carolina 27695.  
E-mail: wlu4@ncsu.edu

<sup>4</sup>Professor, Department of Biostatistics, Columbia University, 722 West 168th Street, New York, NY 10032. E-mail: zj7@cumc.columbia.edu

## A Unified Principal Sufficient Dimension Reduction Package : psdr

*Jungmin Shin*<sup>1</sup>, *Seung Jun Shin*<sup>2</sup>

### Abstract

With the increasing prevalence of high-dimensional data in various academic disciplines, the need for effective dimension reduction techniques in regression analysis has become essential. The psdr package is an R software, which provides a unified framework for principal sufficient dimension reduction (PSDR) methods, encompassing both linear and nonlinear approaches. The package extends the classical linear SDR methods that is functionalized by the R package; 'dr'. Key features of the psdr package include its flexibility in handling various loss functions, even allowing user-defined convex loss functions, as long as they are differentiable everywhere over the support. The software's structure and usage are explained in detail, accompanied by numerical simulations and real data applications to highlight its functionality and utility. The package aims to provide researchers and practitioners with a powerful tool for dimension reduction in high-dimensional regression settings.

*Keywords* : High-dimensional regression, sufficient dimension reduction, R package, various loss functions.

---

<sup>1</sup>Ph.D Candidate Department of the Statistics, Korea University, Anam-ro, Seoungbuk-Gu, 02841, Seoul, Korea. E-mail: jungminshin@korea.ac.kr

<sup>2</sup>(Corresponding Author) Associate Professor, Department of the Statistics, Korea University, Anam-ro, Seoungbuk-Gu, 02841, Seoul, Korea. E-mail: sjshin@korea.ac.kr

## Unpaired Regression for a Discrete Response via Poisson Quantiles Matching

*Hyungjun Lim*<sup>1</sup>, *K. H. Kim*<sup>2</sup>

### Abstract

Analyzing the data collected from different sources requires unpaired data analysis to account for the absence of correspondence between the response variable  $Y$  and explanatory variables  $X$ . Several attempts have been made to analyze continuous  $Y$ , but the response variable of interest may follow a discrete distribution, which previous methodologies have overlooked. To address these limitations, we propose Poisson quantile matching estimation (PQME), the first unpaired data analysis method designed to examine the discrete response variable  $Y$  and the unpaired continuous explanatory variable  $X$ . Using their order statistics, the PQME method matches the linear combination of explanatory variables  $\beta TX$  to  $\ln(Y)$ . An effective algorithm and simulation results are presented, along with the convergence results. We illustrate the practical application of PQME by locating the ideal site for a new facility using real data

*Keywords:* Matching distributions, PQME, Discrete response variable, Unpaired data analysis, Deviance.

---

<sup>1</sup>Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: hlim1410@korea.ac.kr

<sup>2</sup>Department of Statistics, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea.  
E-mail: arlenent@gmail.com

## 절단자료를 위한 의사결정나무

김정환<sup>1</sup>, 조형준<sup>2</sup>

### 요약

의사결정나무는 정해진 의사결정 규칙을 통해 예측하고자 하는 변수를 사각 형태 공간으로 분할하여 분석하는 데이터 마이닝 방법중 하나이다. 주어진 반응변수 공간분할을 반복하면서 하나의 의사결정 나무를 만들고 이를 기반으로 예측을 한다. 이때, 사각 형태로 분류해 나가는 특징으로 인해 의사결정나무는 사용이 쉬우면서 설명력이 뛰어나고 자료가 비선형적이어도 적합하다는 장점을 가지고 있다. 그러나 자료의 특성을 고려하지 않고, 미리 정해진 분류 규칙 기반의 기존 의사결정나무 모형은 데이터의 형태를 고려하여 유연한 분류 규칙을 가진 의사결정나무 모형보다 예측 정확도가 상대적으로 떨어진다는 단점이 있다. 여기서는 자료의 형태가 특정부분에서 관측값이 양의 확률로 다수가 분포되고 다른 부분에서는 연속형의 형태로 분포되어 있거나, 특정한 지점에서 자료가 절단되어 있는 자료에 대해서 유연한 분류규칙을 가지도록 하는 의사결정나무 모형을 제안하였으며, 특히 모형에서 분류변수 선택 시 기존의 전체 탐색방법이 아닌 간차 분석방법을 제안하였다. 모의실험을 통해 모형의 성능을 확인한 결과 연구자가 제안하는 모형이 가장 좋은 성능을 보여주는 것으로 확인되었다.

주요용어 : 절단자료, 의사결정나무, 간차 분석.

<sup>1</sup>02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 박사과정. E-mail: kjhsjy0215@korea.ac.

<sup>2</sup>(교신저자) 02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 교수. E-mail: hj4cho@korea.ac.kr



## 70개 중진료권에 대한 응급실 방문 환자의 유형과 연관성

허영진<sup>1</sup>, 김세형<sup>2</sup>, 이한나<sup>3</sup>, 오미라<sup>4</sup>

### 요 약

2019년 보건복지부에서는 의료 접근성을 높이고 공공의료자원 확충을 위해 전국을 70개 지역으로 구분하였고, 여러 시군구를 포함한 중진료권 단위로 필수의료정책을 관리하기로 하였다. 이에 본 연구는 70개 중진료권별 응급실 방문 환자의 유형과 연관성을 살펴보았다. 2021년 1월 1일부터 12월 31일까지의 국가응급진료정보망(National emergency department information system, NEDIS) 자료에서 3대 중증응급질환(심근경색증, 뇌졸중, 중증외상)으로 내원한 437,349건을 분석하였다. 지역친화도(Relevance index, RI)와 지역환자구성비(Commitment index, CI)에 대한 지역유형은 인접(유출)형 20개 지역과 유입형 50개 지역으로 구분되었다. RI는 CT 평균 대수, MRI 평균 대수와 응급의료기관 수에 대해서 연관성을 보였고, CI는 일차의료중심 의원 수, 재정자주도, 중환자실 수, 입원일 수, 응급의료취약지 수, 평균 진담간호사 수와 연관성을 보였다. 중진료권별 인구, 경제, 자원 등을 이용한 본 연구의 결과들은 응급의료체계의 지역화를 모니터링하고, 보건 의료정책의 수립 등에 활용될 수 있다.

주요용어 : 중진료권, 응급실, 3대 중증응급질환.

---

<sup>1</sup>04564 서울시 중구 을지로 245, 국립중앙의료원 중앙응급의료센터 선임연구원. E-mail: goakzmd@nmc.or.kr  
<sup>2</sup>04564 서울시 중구 을지로 245, 국립중앙의료원 중앙응급의료센터 주임연구원. E-mail: kurukuru5@nmc.or.kr  
<sup>3</sup>04564 서울시 중구 을지로 245, 국립중앙의료원 중앙응급의료센터 주임연구원. E-mail: hanna.lee@nmc.or.kr  
<sup>4</sup>04564 서울시 중구 을지로 245, 국립중앙의료원 중앙응급의료센터 선임연구원. E-mail: omr@nmc.or.kr

## Ensemble Learning for COVID-19 Detection on CT Scan Images

*Van-Phong Truong*<sup>1</sup>, *Thi-Ngot Pham*<sup>2</sup>, *Jun-Ho Huh*<sup>3</sup>, *Chan-Keun Park*<sup>4</sup>

### Abstract

Early identification of COVID-19 may aid in planning a prompt medical response and slowing the deadly disease's fast spread. Recently, a lot of research has been done to analyze medical imaging for the early diagnosis of disease due to the improvement of medical imaging technology as well as the success of deep learning applied to vision tasks. The subject of this work is the use of deep learning based on CT scans to identify COVID-19 patients and severity grading. We utilized ensemble learning for the classification task. The dataset comprises COVID-19 and NORMAL case instances to analyze how well artificial intelligence systems perform at detecting COVID-19 and determining its severity. The outcomes have demonstrated that the ensemble learning method is a potential method to assist radiologists as a thorough, stable, and reliable strategy for early diagnosis of COVID-19.

*Keywords* : COVID-19, Deep Learning, Convolutional Neural Network, CNN, CT Scan image.

### 1. Introduction

The new Coronavirus disease (COVID-19) has been spreading and causing serious problems to life and health of billions of people in the world [1]. The significant awareness of this fatal virus is that it could transmit from person to person with a fast and large-scale spreading via the air and surfaces of infected patients. The two most popular types of images used for COVID-19 detection are Chest X-Ray [2] and CT Scan [3] images. However, the number of patients far exceeds the number of radiologists who take charge of diagnosing the input images. Thus, we need an automatic procedure to help these experts reduce the workload of classifying COVID-19 and NORMAL images (no-covid) [4]. Recently, with the fast development of deep learning techniques applied for medical image analysis, using artificial intelligence deep learning methods to detect COVID-19 for radiology images had been invested in many works [5].

---

<sup>1</sup>Department of Data Informatics, National Korea Maritime and Ocean University, Republic of Korea.  
E-mail: [truongphong2305@g.kmou.ac.kr](mailto:truongphong2305@g.kmou.ac.kr)

<sup>2</sup>(Corresponding Author) Interdisciplinary Major of Ocean Renewable Energy Engineering, National Korea Maritime and Ocean University, Republic of Korea. E-mail: [ngotpham@g.kmou.ac.kr](mailto:ngotpham@g.kmou.ac.kr)

<sup>3</sup>(Corresponding Author) Department of Data Sciences, National Korea Maritime and Ocean University, Republic of Korea. E-mail: [72network@kmou.ac.kr](mailto:72network@kmou.ac.kr)

<sup>4</sup>Department of Data Sciences, National Korea Maritime and Ocean University, Republic of Korea.  
E-mail: [chapark@kmou.ac.kr](mailto:chapark@kmou.ac.kr)

In this paper, we propose an Ensemble model that ensemble learning from several deep learning models to over-performance these separate models (higher confidence) for detecting COVID-19 disease through CT Scan images. We also create an augmented CT Scan dataset with a balanced category (i.e., COVID and NORMAL classes) from three available public datasets, COVID-CT-Dataset, SARS-CoV-2, and COVID-CTSet.

## 2. Methodology

### 2.1. Dataset

To create an image dataset, we collected images from three open source datasets: COVID-CT-Dataset [6], SARS-CoV-2 [7], and COVID-CTSet [8], with a proper image number to have a balanced dataset between two categories (COVID and NORMAL classes). The dataset contribution was summarized in Table 1 with a detailed number of images taken from classes. We divided the dataset with 80% for training and 20% for testing.

### 2.2. Data Augmentation

In Deep Learning (DL), the number of input data plays a crucial role that determining the success or failure of the training model. If we have a rich input dataset with various input images form, we will get the trained model with high generalization capability. Otherwise, if we have very few images, the chance of getting over-fitting is un-avoid. To tackle this problem, data augmentation techniques are applied to boost the model generalization and had worked very well for tasks in which the input data is extremely scarce. In this work, we use five types of image augmentation techniques which widely use in the machine-learning community to enhance our training input images. These techniques include height\_shift\_range, rotation\_range, width\_shift\_range, brightness\_range, and zoom\_range as shown in Figure 1.

Table 1. COVID CT Scan images contribution from classes.

Dataset	Train		Test	
	COVID	NORMAL	COVID	NORMAL
	1282	1301	319	325

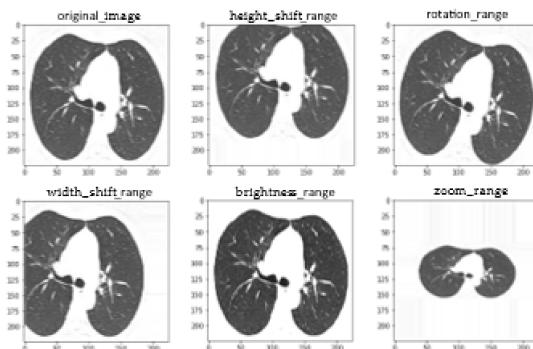


Figure 1. Images augmentation techniques used for experiments

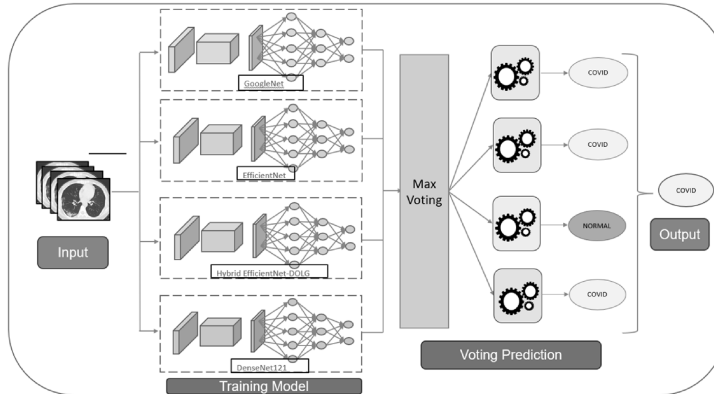


Figure 2. Architecture of COVID CT Scan classification using ensemble learning

### 2.3. Model Architectures

In this paper, we utilized ensemble learning for the COVID-19 CT scan image classification task. Ensemble learning is used for classification tasks to make a prediction based on voting between multiple models. Our experiment used max voting as the voting method for the final prediction as shown in Figure 2. Instead of training each model separately, the voting classifier makes decisions concerning the highest primary vote class. We used four deep-learning models to run our experiments on COVID-19 CT image dataset. These models we use include, GoogleNet [9], EfficientNet [10], Hybrid EfficientNet-DOLG [11], and DenseNet121 [12] models.

## 3. Experimental Results and Conclusion

In our experiment, we trained deep learning models on the computer hardware setting: CPU Intel(R) Xeon(R) @ 2.20GHz RAM 32GB. We utilized Tensorflow Keras as the implementation library for deep learning models. We use standard convolutional neural networks from the library and fine-tune them for our experiments. We use ImageNet pre-trained weights for each architecture and set LearningRateScheduler, EarlyStopping, ModelCheckpoint, ReduceOnPlateau based on Tensor Flow Keras built-in API.

LearningRateScheduler gets a new learning rate at the beginning of each epoch and applies this learning rate to the optimizer. EarlyStopping monitored the loss at the end of each epoch to evaluate the increase and decrease of the metric, if there is no improvement in the monitored metric, the training will be terminated. ModelCheckpoint saves models or weights at some training interval, then saved models or weights can be loaded and continued to train. ReduceOnPlateau helps to reduce the learning rate when loss increase, and the metric does not improve. Based on the number of patience if there is no improvement, the learning rate will be lessened.

Table 2. Covid detection performance of five Deep Learning methods based on a precision metric.

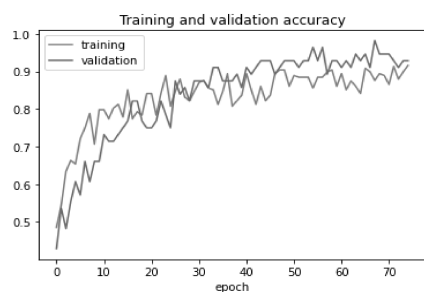
Method	Precision Metric			
	COVID	NORMAL	Macro Avg	Micro Avg
GoogleNet	0.77	0.75	0.76	0.76
EfficientNet	0.85	0.72	0.79	0.79
Hybrid EfficientNet-DOLG	0.85	0.81	0.83	0.83
DenseNet121	0.92	0.84	0.88	0.88
Proposed Ensemble	<b>0.94</b>	<b>0.88</b>	<b>0.89</b>	<b>0.90</b>

Table 3. Covid detection performance of five Deep Learning methods based on recall metric.

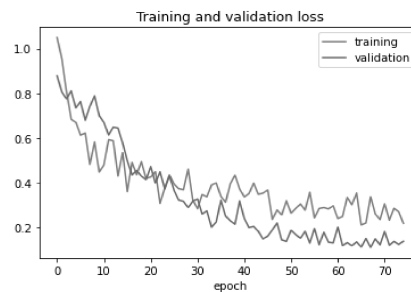
Method	Recall Metric			
	COVID	NORMAL	Macro Avg	Micro Avg
GoogleNet	0.73	0.78	0.76	0.76
EfficientNet	0.66	0.89	0.77	0.77
Hybrid EfficientNet-DOLG	0.80	0.86	0.83	0.83
DenseNet121	0.82	0.93	0.87	0.88
Proposed Ensemble	<b>0.88</b>	<b>0.96</b>	<b>0.90</b>	<b>0.92</b>

Table 4. Covid detection performance of five Deep Learning methods based on F1-score metric.

Model	F1-score Metric			
	COVID	NORMAL	Macro Avg	Micro Avg
GoogleNet	0.75	0.76	0.75	0.75
EfficientNet	0.74	0.80	0.77	0.77
Hybrid EfficientNet-DOLG	0.82	0.83	0.83	0.83
DenseNet121	0.87	0.88	0.87	0.87
Proposed Ensemble	<b>0.90</b>	<b>0.93</b>	<b>0.92</b>	<b>0.88</b>



(a) Training and validation accuracy



(b) Training and validation loss

Figure 3. Training and validation Accuracy, Loss of DenseNet121.

From Tables 2, 3, and 4, we can see that the Ensemble model produces the highest results on both precisions, recall, and f1-score metrics, and next after is the Dense121 model shown on Figure 3. Compare to each model which was trained separately, ensemble learning shows outperformed results. These numerical results prove that deep learning methods could apply to medical image classification with reliability and high performance. Specifically, in this paper, artificial neural networks can be used as the computer aid diagnosis for COVID-19 detection.

In this work, we study the application of deep learning for early COVID-19 detection using CT Scan images. Ensemble Learning from four deep learning models is utilized to form the outstanding model for COVID-19 CT Scan images. We create an augmented COVID-19 CT Scan dataset with a balanced and clean training and testing set. This dataset is run on the four deep learning models and then ensemble learning with our proposed ensemble model. The output results indicated that the proposed ensemble model has higher confidence in detecting COVID-19 patients using CT Scan images.

## References

- Wu, F.; et al, "A new coronavirus associated with human respiratory disease in China," *Nature*, Vol. 579, pp. 256-269, 2020.
- Jacobi, A.; Chung, M.; Bernheim, A.; Eber, C., "Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review," *Clinical imaging, Elsevier*, Vol. 64, pp. 35-42, 2020.
- Li, Y.; Xia, L., "Oronavirus Disease 2019 (COVID-19): Role of Chest CT in Diagnosis and Management," *Ajr. Am. J. Roentgenol, American Roentgen Ray Society*, Vol. 214, pp. 1280-1286, 2019.
- Khanna, M.; Agarwal, A.; Singh, L.K.; Thawkar, S.; Khanna, A.; Gupta, D., "Radiologist-level two novel and robust automated computer-aided prediction models for early detection of COVID-19 infection from chest X-ray images," *Arabian Journal for Science and Engineering, Springer*, Vol. 7, pp. 1-33, 2021.
- Bram van Ginneken, "The Potential of Artificial Intelligence to Analyze Chest Radiographs for Signs of COVID-19 Pneumonia," *Radiology, RSNA*, Vol. 299, pp. 1-2, 2021.
- Zhao, Jinyu.; Yichen Zhang.; Xuehai He.; Pengtao Xie., "COVID-CT-dataset: a CT scan dataset about COVID-19," *arXiv preprint arXiv:2003.13865* 490, 2020.
- Angelov.; Plamen.; Eduardo Almeida Soares, "SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification," *MedRxiv*, 2020.
- Rahimzadeh, M., Attar, A., Sakhaei, S. M, "A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset," *Biomedical Signal Processing and Control, Elsevier*, Vol. 68, p. 102588, 2021.
- Szegedy, Christian, et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition, CVPR, IEEE*, pp. 1-9, 2015.
- Tan, Mingxing, Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning, PMLR*, pp. 6105-6114, 2019.
- C. Henkel, "Efficient large-scale image retrieval with deep feature orthogonality and Hybrid-Swin-Transformers," *arXiv preprint arXiv:2110.03786*, 2021.
- Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger., "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition, CVPR, IEEE*, pp. 4700-4708, 2017.
- Thi-Ngot Pham, Viet-Hoan Nguyen, Jun-Ho Huh. "Integration of improved YOLOv5 for face mask detector and auto-labeling to generate dataset for fighting against COVID-19," *The Journal of Supercomputing, Springer*, pp. 1-27, 2023.

## COVID-19 팬데믹 상황에서 간호대학생의 온라인 취업적응프로그램의 효과

유제복<sup>1</sup>

### 요 약

본 연구는 코비드19의 팬데믹 상황에서 간호대학생의 온라인 취업적응프로그램이 그릿, 진로결정자기효능감, 취업스트레스, 임상수행능력에 미치는 영향을 확인하고자 K도 J시 소재 K대학의 재학생 34명을 대상으로 자료를 수집하였다. 간호대학생에게 온라인 취업적응프로그램을 개발하여 1일 4시간씩, 5일 동안 10회 온라인 교육을 실시하였다. 수집된 자료는 기술통계량과 회귀분석을 이용하였다. 연구결과 그릿( $F=17.789, p=.000$ ), 진로결정자기효능감( $F=22.472, p=.000$ ), 임상수행능력( $F=19.473, p=.000$ )은 대상자가 온라인 취업적응프로그램에 참여 전, 참여 후, 1달 후의 평균값이 증가하였고, 시간이 지날수록 유의미한 차이를 보였다. 본 연구의 결과를 토대로 간호대학생의 다양한 취업적응프로그램의 개발과 적용을 제언하는 바이다.

주요용어 : COVID-19 팬데믹, 간호대학생, 온라인 취업적응프로그램.

### 1. 서론

#### 1.1. 연구의 필요성과 목적

최근 신규간호사의 높은 조기 이직률은 환자안전과 의료서비스 질 저하로 사회적인 관심사로 부상하면서 간호교육현장에서 간호대학생의 취업 전 체계적인 취업적응프로그램 개발 및 습득에 대한 요구 높아지고 있다(Jo, 2019). 간호대학생은 끊임없이 변화하는 임상현장에서 간호사로서의 수행할 수 있는 역량이 필요하나 정규교육과목으로서 취업적응프로그램이 매우 부족하며, 또한 졸업 전 첫 직장 선택을 어떻게 준비해야 하는지에 대한 준비없이 취업을 하고 있는 실정이다(Park, Lee, 2016).

2020년 초 전 세계를 강타한 COVID-19 팬데믹은 갑작스럽게 교육 운영의 변화가 있었으며, 2017년 4월 통계청의 고용동향에 따르면 청년실업률은 11.2%로 전체 실업률 4.2%보다 2배 이상 높으며(Beak, 2020), 간호대학생의 경우 취업률은 높지만 1년 이내의 이직률(병원간호사회, 2022) 또한 높아서 국가적으로 간호사 부족 현상과 교육 투자 대비 손실이 큰 상황이 점점 악화되고 있다.

그렇지만, 목표를 달성하기 위한 장기적인 열정과 인내를 의미하며, 핵심 속성은 흥미의 유지와 노력의 지속성이다(Jeong, Jung, 2018). 간호대학생은 미래의 간호사로 졸업 후에 간호사가 되어 병원에서 신체적·정신적 어려움을 경험하게 될 때 참고 견디며, 장기적인 목표 달성을 위해 꾸준한 노력과 지속성이 필요하다(Jeong, Jung, 2018). 진로결정자기효능감은 개인이 진로결과와 관련된 과제들을 성공적으로 해결할 수 있는 자신감 혹은 진로 선택과정에서 최선의 결정을 내릴 수 있다는

<sup>1</sup>52725 경남 진주시 동진로 33, 경상국립대학교 간호학과 부교수. E-mail: jbyoo@gnu.ac.kr

확신감의 정도이다(Taylor, Betz, 1983). 따라서 진로결정자기효능감이 높은 간호대학생일수록 진로 탐색의 의도가 높고 그 결과 올바른 진로를 결정을 내린 후 그것을 잘 실행할 가능성이 높다고 하였다. 취업스트레스란 개인의 구직 과정에서 취업 준비 혹은 실패 등으로 인해 발생하게 되는 모든 스트레스를 일컫는다(Hong, 2015). 경쟁력을 갖추어 조금 더 나은 곳에 취업을 하기 위한 간호대학생에게도 취업스트레스는 예외일 수 없고(Ahn, Kim, 2018), 취업 경쟁률과 스트레스는 더욱 심화되고 있다(Seo, Lee, 2015). 임상수행능력은 업무수행을 위한 지식, 기술, 태도의 통합된 능력의 집합체로 간호역할에 대한 요구를 효과적으로 감당할 수 있는 능력이다(Campbell, Mackay, 2001). 특히 COVID-19 팬데믹 상황의 일선에서 국민건강을 위해 역량을 발휘하는 간호사들의 활약은 많은 국민들에게 간호사의 중요성을 새삼 일깨워주는 계기가 되기도 하였고(Kim, Yu, 2121), 이러한 재난 상황에서 간호사, 그리고 간호학생들의 임상수행능력은 그 어느 때보다 중요성이 더 증가하였고 신규간호사의 적응을 도와줄 수 있는 중요한 전략으로 사료된다.

본 연구는 COVID-19 팬데믹 상황이 장기화되어 대면교육이 제한됨에 따라 학생들의 취업능력강화를 위해 온라인 취업능력강화프로그램을 제공하고, 프로그램 제공 전·후의 효과를 확인하여 현재의 상황을 파악하고 향후 대책을 마련하고자 진행하였다. 그러나 국내에서 간호대학생을 대상으로 온라인으로 개발된 학생의 자기주도적 취업적응 교육프로그램 개발은 매우 제한적이다.

이에 본 연구에서는 간호대학생의 온라인 취업적응교육프로그램을 개발하고 적용하여 간호대학생의 그릿, 진로결정자기효능감 및 취업스트레스와 임상수행능력에 미치는 효과를 확인하고 취업적응능력 함양을 위한 전략을 수립하는데 기초자료로 활용하고자 한다.

## 1.2. 연구의 목적

본 연구는 간호대학생을 대상으로 취업적응프로그램이 그릿과 진로결정자기효능감, 취업스트레스, 임상수행능력에 미치는 효과를 파악하기 위함이다.

## 2. 연구방법

### 2.1. 연구설계

본 연구는 COVID-19 팬데믹 상황에서 간호대학생의 온라인 취업적응프로그램이 그릿과 진로결정효능감, 취업스트레스와 임상수행능력에 미치는 영향을 파악하기 위한 단일군 전, 후 유사실험연구이다.

### 2.2. 연구대상 및 자료수집방법

본 연구는 K도 J시 소재 K대학의 간호학과 재학생으로, 연구 참여에 동의한 자를 대상으로 하였다. G-power 3.1.9.4 version을 이용하여, 선행연구(Jeon, 2020)의 결과를 근거로 RMANOVA에 필요한 유의수준  $\alpha=.05$ , 효과크기.25(medium), 검정력(Power=.80) 최소 표본크기는 34명의 자료를 분석하였다. 본 연구는 IRB심의위원회의 승인(제CSIRB-2022002)후 자료를 수집하였다.



### 2.3. 연구변수

본 연구는 COVID-19 팬데믹 상황에 간호대학생에게 1일 4시간씩 5일동안 10회차의 온라인취업 적응 교육프로그램을 적용하고 일반적인 특성 5문항, 그릿 12문항, 진로결정자기효능감 25문항, 취업스트레스 22문항과 임상수행능력 7문항으로 총 71문항의 구조화 된 설문지를 구성하였고, 5점 Likert 척도로 각 연구 도구들은 저자에게 문자와 e-mail을 통하여 사전 허락을 받았다. 그릿은 Original Grit Scale(Grit-O)을 Lee(2015)가 변안 도구를, 진로결정자기효능감은 Career Decision-Making Self-Efficacy Scale 단축형 척도를 Lee(2001)의 변안한 도구를, 취업스트레스는 코넬 대학 ‘스트레스 측정법’(Cornell Medical Index)을 Kang(2006)이 수정·보완한 도구를, 임상수행능력은 간호학 학습성과 간접측정 도구를 임상실무통합역량도구로 수정·보완 한 Kim(2015)의 도구를 이용하였다.

### 2.4. 자료분석방법

수집된 자료는 SAS 9.4를 이용하여 실수, 백분율로 연속변수의 경우 평균과 표준편차와 반복측정분산분석(repeated measure ANOVA; RMANOVA)을 이용하였다.

## 3. 연구결과

### 3.1. 대상자의 일반적 특성

대상자의 연령은 평균  $21.77 \pm 1.17$ 세이었으며, 성별은 여자 22명(64.7%)이었고, 2학년이 4명(11.8%), 3학년이 18명(52.9%), 4학년이 12명(35.3%) 이었다. 간호학과 지원 동기는 적성에 맞아서가 5명(14.7%), 취업을 위해서가 9명(26.5%), 다양한 업종에 취업이 가능해서가 11명(32.4%)이었고 선호하는 직장은 대학병원이 31명(91.2%) 이었다(Table 1).

Table 1. General characteristics of subjects (N=34)

Variables	Categories	n	%	Mean±SD	Range
Age(yr)	19	7	20.6	21.77±1.17	19-23
	20	14	41.2		
	21	3	8.8		
	22	5	14.7		
	23	5	14.7		
Gender	Female	22	64.7		
	Male	12	35.3		
Grade	Sophomore	4	11.8		
	Junior	18	52.9		
	Senior	12	35.3		
Reason for application to nursing school	Fit of aptitude	5	14.7		
	Beneficial to find a job	9	26.5		
	Valuable work	11	32.4		
	Steady job	6	17.6		
	On the advice of parents or others	3	8.8		
Preferred workplace	University hospital	31	91.2		
	General hospital	3	8.8		

Table 2. Scores of subject's online employment adaptation program (N=34)

Scale	Domain	pre test Mean±SD	post test 1 Mean±SD	post test 2 Mean±SD	F	p
Grit		3.28±.37	3.44±.36	3.50±.37	6.49	.003*
Career decision-making self-efficiency		3.47±.53	3.78±.59	3.78±.69	7.42	.001*
Employment stress		1.91±.61	1.99±.69	2.05±.82	0.82	.447
	Total	3.19±.58	3.69±.58	3.80±.59	16.95	.000*
Clinical nursing practice competency	medication	3.35±.69	4.41±.66	4.15±.50	32.94	.000*
	bed sore	3.53±.71	4.38±.70	4.35±.65	26.23	.000*
	transfusion	3.35±.85	4.24±.61	4.21±.59	21.47	.000*
	infection	3.21±.59	4.38±.65	4.26±.62	48.44	.000*

### 3.2. 대상자의 온라인 취업적응프로그램의 효과

간호대학생은 온라인 취업교육 프로그램 적용 전·후·1달 후의 그릿(F=6.49, p=.003), 진로결정자기 효능감(F=.742, p=.001), 임상수행능력(F=16.95, P=.000)은 시간이 지남에 따라 유의하게(F=6.49, 증가하였다(Table 2). 임상수행능력 중 투약간호는(F=32.94, p=.000), 욕창간호는(F=26.23, p=.000), 수혈간호는 (F=21.47, p=.000), 감염관리는 (F=48.44, p=.000)으로 평균점수가 증가하였고, 유의하였다 (Table 2).

## 5. 결론 및 제언

본 연구결과는 간호대학생이 온라인 취업교육프로그램 적용 전, 후, 1달 후 그릿과 진로결정자기 효능감, 취업스트레스와 임상수행능력의 효과를 확인하여 향후 간호대학생의 비교과 교육과정 등 다양한 취업교육프로그램 개발의 기초자료를 제공하고자 한다. 본 연구의 결과는 자료수집의 편의성을 위해 일부 지역의 간호대학생을 대상으로 설문지를 이용하여 자료를 수집하였으므로, 결과에 대한 일반화에 제한점이 있으므로 추후 연구에서는 대조군이 있는 반복연구를 통해 장기 효과를 확인 해 볼 것을 제언하는 바이다.

## References

- Kim, Y. M. (2022). The influence of grit and organizational socialization on clinical competency in new graduate nurses, Master's thesis, Ajou University. Gyeonggi-Do. (in Korean)
- Seo, B. J. (2021). The mediating effect of career decision-making self-efficacy in the relationship between major satisfaction and career preparation behavior in university students majoring in social welfare, *Journal of learner-Centered Curriculum and Instruction*, 21(19), 217-230. <http://doi.org/10.22251/jlcci.2021.21.19.217>
- Yoo, J. B., Kim, J. S. (2022). The effect of grit, nursing practical proficiency and clinical competency in perianesthesia nurses, *Journal of the Korean data Analysis Society*, 24(6), 2355-2368.(in Korean). DOI: <http://doi.org/10.37727/jkdas.2022.24.6.2355>
- Won, J. S., Yoo, J. B. (2018). Impact of job stress, organizational culture and professional identity on job satisfaction of post-anesthesia care unit nurses, *Journal of the Korean data Analysis Society*, 20(6), 3211-3325.(in Korean). DOI: [10.37727/jkdas.2018.20.6.3211](http://doi.org/10.37727/jkdas.2018.20.6.3211)

## CEO Pay-performance Sensitivities in the ESG Managed Firms<sup>\*</sup>

*Injoong Kim*<sup>1</sup>

### Abstract

CEOs of highly rated ESG firms are paid significantly higher salaries. High ESG firms are larger, more profitable, and mature firms, facing less risks compared to low ESG rated firms. Although these systematic differences in firm characteristics account for a significant portion of the CEO pays, when we decompose the CEO pays between different ESG groups, there exists a non-trivial residual portion of the unexplained compensation gap that the traditional firm characteristics can not fully explain. About 79.8% of the mean difference is attributable to the firm characteristic part and the residual 20.2% can be ascribed to the ESG group effect. The high ESG firms are more active in using the incentive compensation scheme to discipline managers while the low ESG firms are more conservative in adopting the sensitive incentive pay systems because they are more susceptible to the negative factors such as risk and leverage. As a result, we observe significantly higher pay-performance sensitivities for the high ESG group and the low ESG firms are paid lower than what their firm characteristics prescribe. The ESG rating has an orthogonal part of the explanatory power that contributes to the pay-performance sensitivity over the traditional solvency measure such as credit ratings.

*Keywords* : CEO pay, compensation, performance, sensitivity, ESG.

### 1. Introduction

The incentive pay contract of the Chief Executive Officer (CEO) contributes to motivating managerial efforts and reducing the agency problem. In this paper, we examine the CEO pays of ESG managed firms and test the marginal impact of the ESG management on the pay performance sensitivity. Using the decomposition method of Jann (2008), we decompose the mean difference of the CEO payments between different ESG groups into the part that is due to the difference in the endowments of firm characteristics and the residual component that is attributable to the ESG group effect.

---

<sup>\*</sup>This work was supported by 2023 Hongik University research fund.

<sup>1</sup>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

We analyze KOSPI and KOSDAQ firms that have ESG rating information in the KCGS website. The mean difference of the average CEO pays between the two ESG groups A and B can be decomposed as follows.

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

$$E[Y_A] - E[Y_B] = \{E[X_A] - E[X_B]\}\beta_P + \{E[X_A](\beta_A - \beta_P) + E[X_B](\beta_P - \beta_B)\} \quad (2)$$

In all specifications, the first component commonly represents the portion of the pay difference that is explained by the group difference in the endowment vector of firm characteristics, and the residual terms are the group difference that is not explained by the firm characteristics. Based on the framework of Gao, Li (2015), we estimate the following panel data regressions and use them as an input for the decomposition.

$$\begin{aligned} \ln(CEO\ Pay) = & \alpha + \beta_1 Ind\_ESG + \beta_2 \ln(total\ assets) + \beta_3 Performance\ measures \\ & + \beta_4 Other\ firm\ characteristics + Industry\ FEs + Year\ FEs + \epsilon \end{aligned} \quad (3)$$

## 3. Empirical Results

Table 1 reports the average values of firm characteristic variables for both the high and the low ESG rated firms. The CEOs of high ESG firms are paid more than those of low ESG firms. High ESG firms are generally larger and more profitable. They are at their mature stage of firm lifecycle with more accumulated retained earnings over the total equity. On the other hand, low ESG firms are not as profitable and they are generally younger and face greater risks.

Table 1. Means of firm characteristic variables for different ESG rated firms

	ln(pay)	Size	ROA	RE/TE	Leverages	CF volatility	Risk
High ESG	13.366	21.932	0.047	0.569	0.437	0.018	3.321
Low ESG	12.414	19.714	0.030	0.430	0.389	0.017	3.862
Total	12.624	20.202	0.034	0.461	0.399	0.017	3.743
High - Low	0.951* (<.001)	2.217* (<.001)	0.018* (<.001)	0.139* (<.001)	0.048* (<.001)	0.001 (0.093)	-0.541* (<.001)

Note: Table 1 summarizes the averages of key variables in each ESG group for the entire sample period. On the bottom, the reported are the T-tests of mean difference between the two ESG groups. Numbers in parentheses are the standard errors. \* indicates the mean difference is statistically significant at the 1% level.

Table 2. Decomposition of CEO compensations

	Threefold					Twofold	
	Ln(Pay)	Exponentiate d	Ln(Pay)	Exponentiate d		Ln(Pay)	Exponentiate d
Low ESG	12.414	246,344.8	12,414	246,344.8	Low ESG	12,414	246,344.8
High ESG	13.366	637,811.1	13.366	637,811.1	High ESG	13.366	637,811.1
Difference	-0.951* (0.026)	0.386* (0.010)	-0.951* (0.026)	0.386* (0.010)	Difference	-0.951* (0.026)	0.386* (0.010)
Endowments	-0.759* (0.036)	0.468* (0.017)	-0.761* (0.023)	0.467* (0.011)	Explained	-0.758* (0.023)	0.469* (0.011)
Coefficients	-0.191* (0.029)	0.826* (0.024)	-0.193* (0.040)	0.825* (0.033)	Unexplained	-0.193* (0.027)	0.824* (0.022)
Interaction	-0.002 (0.038)	0.998 (0.038)	0.002 (0.038)	1.002 (0.038)			

Note: Numbers in parentheses are the standard errors, and \* indicates the statistical significance at 1% level.

Table 3. Factors influencing CEO pay-performance sensitivity

	(1)	(2)	(3)	(4)
Risk × ROA	-0.064* (0.015)			
Ind_ESG × ROA	0.654* (0.243)	0.670* (0.244)	0.765* (0.253)	0.727* (0.245)
Credit rating			-0.044* (0.008)	-0.042* (0.008)
Credit rating × ROA				-0.096 (0.055)
Size	0.319* (0.008)	0.319* (0.008)	0.318* (0.008)	0.318* (0.008)
ROA	1.229* (0.167)	1.065* (0.155)	1.047* (0.176)	1.106* (0.229)
CF volatility	7.173* (0.641)	7.324* (0.641)	7.170* (0.640)	6.582* (0.786)
Cash	0.389* (0.129)	0.373* (0.129)	0.341* (0.128)	0.344* (0.129)
Leverage	-0.081* (0.029)	-0.092* (0.031)	-0.059* (0.023)	-0.058* (0.022)
RE/TE	0.120* (0.013)	0.105* (0.012)	0.094* (0.012)	0.096* (0.011)
Risk	-0.023* (0.007)	-0.024* (0.007)	-0.022* (0.008)	-0.021* (0.008)
Foreign share	0.345* (0.078)	0.355* (0.078)	0.315* (0.077)	0.308* (0.078)
Ind_ESG	0.151* (0.031)	0.152* (0.030)	0.142* (0.030)	0.144* (0.031)
Constant	5.537* (0.262)	5.551* (0.263)	5.724* (0.264)	5.715* (0.263)
Adj. R2	0.447	0.446	0.449	0.449

Note: Numbers in parentheses are the standard errors, and \* indicates the significance at 1% level.

The Table 2 decomposes the mean difference. Major difference in the CEO pays exists between the high ESG and the low ESG group. In percentage terms, roughly 80% (0.759/0.951) is due to the endowments and the residual 20% is ascribed to the ESG group effect. We observe non-trivial ESG effect and there exists a clear difference in CEO pays between different ESG rated firms that can not be fully explained by firm characteristics. Table 3 investigates factors that influence the CEO pay-performance sensitivity by incorporating various interaction terms.  $\text{Ind\_ESG} \times \text{ROA}$  can capture the incremental difference in the pay-performance sensitivity between the two different ESG groups. In all specifications of the model, the coefficients on the ROA performance measure is significant supporting that the CEO pay is largely responsive to the firm performance. In addition, the significance of the interaction terms suggests that the high ESG firms are more actively using the incentive compensation schemes to discipline managers while the low ESG firm's pay is not as sensitive.

## 5. Conclusions

In this paper, we analyze the CEO compensation for the ESG participating firms and investigate factors that influence the CEO pay performance sensitivity. When we decompose the mean difference in the CEO pays between different ESG groups, there exists a residual portion of the unexplained compensation gap that the traditional firm characteristic measures can not sufficiently explain. We find evidence that the high ESG firms are more active in using the incentive compensation schemes to discipline managers while the low ESG firms are more conservative in adopting the sensitive incentive pay systems.

## 호텔·외식 고객의 기술불안과 행동의도 간의 영향관계: 셀프서비스기술 서비스품질과 고객만족의 다중매개효과\*

문혜영<sup>1</sup>

### 요 약

본 연구는 호텔·외식산업 셀프서비스기술 이용고객의 기술불안, 셀프서비스기술 서비스품질(SSTQUAL), 고객만족 및 행동의도 간의 구조적 관계를 살펴보고, 또한 SSTQUAL 및 고객만족의 다중매개효과를 검증하고자 하였다. 설문조사는 호텔·외식기업 SST를 이용한 경험이 있는 성인(18세 이상) 고객을 대상으로 진행되었으며, 유효표본 282부를 활용하여 분석을 수행하였다. 본 연구의 분석결과는 다음과 같다. 첫째, 기술불안은 SSTQUAL 요인( $\beta=-.389, p<.01$ ), 행동의도( $\beta=-.136, p<.01$ )에 부정적인 영향을 미치는 것으로 나타났으나, 고객만족에는 유의한 영향을 미치지 않는 것으로 나타났다( $\beta=-.053, p>.05$ ). 둘째, SSTQUAL 요인은 고객만족( $\beta=.826, p<.01$ ), 행동의도( $\beta=.271, p<.01$ )에 긍정적인 영향을 미치는 것으로 나타났다. 셋째, 고객만족은 행동의도에 정(+의 영향을 미치는 것으로 나타났다( $\beta=.521, p<.01$ ). 마지막으로 고객의 기술불안과 행동의도 간의 영향관계에서 SSTQUAL, 고객만족의 다중매개효과는 기술불안-SSTQUAL-행동의도 경로에서 유의적으로 나타났다. 따라서 호텔·외식 고객의 SST 이용에 대한 행동의도를 높이기 위해서는 기술불안감을 낮춰야 하며, 기능, 고객화, 디자인 등의 SSTQUAL 관리가 중요시 된다.

주요용어 : 기술불안, 셀프서비스기술 서비스품질(SSTQUAL), 고객만족, 행동의도, 호텔·외식산업.

### 1. 서론

셀프서비스기술(self-service technology: SST) 인터페이스는 비즈니스의 장기적인 성공에 의미 있는 영향을 미치며, 인력의 효율적 활용을 가능하게 한다고 한다(Shin, Perdue, 2019). 고객측면에서는 전통적인 인적서비스 대신 편리성을 제공하며(Safaeimanesh, Kılıç, Alipour, Safaeimanesh, 2021), 서비스에 대한 상당한 통제력이 고객에게 주어지기도 한다(Lyu, Lim, Choi, 2019). 이러한 기술주도 서비스 환경에서의 고객이 지각하는 서비스품질 평가연구는 필수적인 과제가 되고 있으며, 고객-직원간 대상으로 개발된 SERVQUAL(Parasuraman et al., 1988)은 전통적 시장과 커다란 차이가 존재하는 셀프서비스기술기반 서비스 품질을 측정하는 데에 많은 한계가 있다는 것이 지적되어 왔다(Lin, Hsieh, 2011). Dabholkar(1996)의 연구를 시작으로 SST 분야의 서비스품질에 대한 연구는 Lin,

\* 본 논문은 또는 저서는 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A5B5A16082291).

<sup>1</sup>03722 서울시 서대문구 연세로 50, 연세대학교 심바이오틱 라이프텍 연구소 연구교수.  
E-mail: hye-moon@hanmail.net

Hsieh(2011)가 다양한 서비스 산업을 대상으로 셀프서비스기술 서비스품질(SSTQUAL) 평가지표를 개발하였으며, 후속학자들은 SSTUQUAL 평가지표를 기반으로 산업특성 및 SST 유형을 고려하여 수정·변형하여 사용해 오고 있다. 외식산업의 경우 패스트푸드를 시작으로 키오스크 및 모바일 기반 SST를 폭넓게 적용하고 있으며 호텔기업도 COVID-19 팬데믹 기간동안 빠르게 다양한 서비스 인카운터에 SST를 도입·활용하고 있다. 고객-기술 상호작용 연구에서도 서비스품질과 고객만족 간의 영향관계가 중요한 의미가 있음이 밝혀지고 있으며(Ghosh, 2020; Ribbink, Riel, Liljander, Streukens, 2004), Gelbrich, Sattler(2014)는 낯선 사람들이 있는 공공장소에서 개인이 갖는 기술 불안감은 SST에 대한 사용의도에 부정적인 영향을 미치며, 지각된 시간압박과 함께 지각된 혼잡도가 결합이 될 때 SST에 대한 사용의도는 완전히 억제될 수 있다고 한다. 따라서 전형적인 인적기반의 서비스 산업이었던 호텔·외식 기업에서 SST 도입을 통해 기술기반 서비스 인카운터로의 전환이 시도되고 있는 시점에서 고객이 지각하고 있는 기술불안과 SST에 대한 행동의도 간의 영향관계에서 SSTUQUAL과 고객만족의 다중매개효과를 검증하고자 하였다.

## 2. 연구설계

### 2.1. 연구모형

호텔·외식산업 셀프서비스기술 이용고객의 기술불안, SSTQUAL, 고객만족 및 행동의도 간의 구조적 영향관계를 살펴보기 위해 Figure 1의 연구모형(부분매개모형)을 설정하였다.

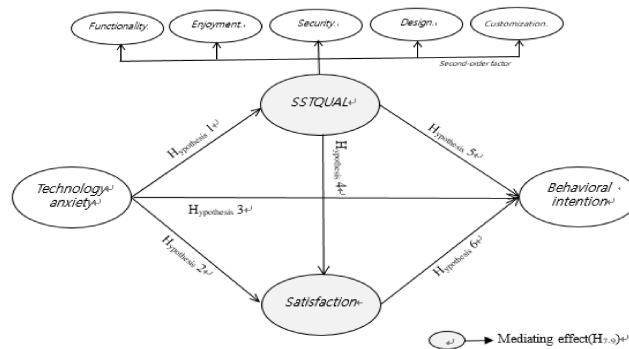


Figure 1. Research model

### 2.2. 조사설계

본 연구의 설문조사는 최근 3개월 이내 호텔 및 외식기업의 셀프서비스기술(SST)을 이용한 경험이 있는 성인(18세 이상) 고객을 대상으로 진행되었으며, 조사방식은 1차적으로 모집문건(목적, 대상, 기간, 소요시간 및 보상/기프티콘 제공)을 온라인 사이트에 올려 대상자를 수집하였으며(URL 접속을 통하여 조사동의 후 설문응답진행), 2차적으로 편의표본추출방식 중 눈덩이 표본추출(snowball sampling)을 통해 SNS로 구글 설문지가 전달되었다. 총 295부가 수집되었으며, 불성실한 데이터를 제거하고 282부의 유효 표본이 분석에 활용되었고, 통계적 처리는 SPSS 및 AMOS를 통해 빈도분석, 신뢰도 및 상관분석, 확인요인분석, 구조방정식모형(SEM)을 수행하였다.



### 3. 연구결과

#### 3.1. 표본의 인구통계적 특성

표본 282명의 성별, 연령, 결혼여부, 학력, 직업 등의 인구통계적 특성에 대해 빈도분석을 실시한 결과, 성별은 남성이 84명(29.8%), 여성이 198명(70.2%)으로 나타났고, 연령은 20대가 83명(29.4%), 30대가 92명(32.6%), 40대가 74명(26.2%), 50대 이상이 33명(11.7%)로 나타났다. 결혼여부는 기혼이 51.8%, 미혼이 47.5%로 비슷한 수준으로 나타났으며, 학력은 대학 및 대학교 졸업이 175명(62.9%), 대학원 졸업 이상이 104명(36.9%)을 차지하였다.

#### 3.2. 측정모형의 타당성 및 신뢰성 평가

연구모형의 적합도를 판단하기 위해 이차원 확인요인분석을 실시하였으며, 측정모형 적합도 지수는  $\chi^2=680.773$ ,  $df=313$ ,  $\chi^2/df=2.175$ ,  $RMSEA=.065$ ,  $NFI=.884$ ,  $IFI=.934$ ,  $TLI=.925$ ,  $CFI=.933$  으로 나타나 전체적으로 적합도 기준치를 충족시키고 있다. 또한 집중타당성을 평가하기 위해 개념신뢰도(CR)를 살펴본 결과, .849-.948로 나타나 기준치인 .7이상(Fornell, Larcker, 1981)을 확보하였으며, 평균분산추출값(AVE)도 .585-.820으로 기준치인 .5이상(Hair, Black, Babin, Anderson, 2010)을 나타내어 집중타당성을 확보하였다. 이 연구의 주요 변수에 대한 신뢰도 분석을 실시한 결과, Cronbach's alpha값이 .868~.930으로 기준치인 .6 이상(Nunnally, 1978)으로 나타나 높은 신뢰수준을 확보하였다. 또한 판별타당성 검증을 위해 구성개념 간의 상관계수의 제곱값( $r^2$ )과 평균분산추출값을 비교한 결과, 상관계수 제곱값( $r^2$ )보다 평균분산추출값이 큰 것(.585-.820)로 나타나 판별타당성이 확보되었다.

#### 3.3. 가설 검증

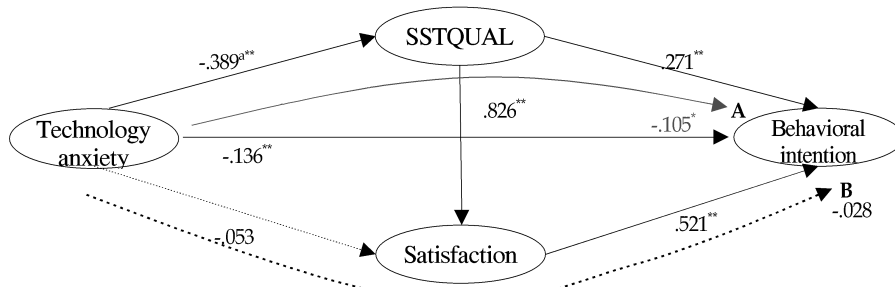
최종 구조모형 적합도 지수는  $\chi^2=680.773$ ,  $df=313$ ,  $\chi^2/df=2.175$ ,  $RMSEA=.065$ ,  $NFI=.884$ ,  $IFI=.934$ ,  $TLI=.925$ ,  $CFI=.933$  등으로 나타나 전체적으로 적합도 기준치를 충족시키고 있다(Table 1). 경로별 결과는 기술불안은 SSTQUAL 요인( $\beta=-.389$ ,  $p<.01$ ), 행동의도( $\beta=-.136$ ,  $p<.01$ )에 부정적인 영향을 미치는 것으로 나타났으나, 고객만족에는 유의한 영향을 미치지 않는 것으로 나타났다( $\beta=-.053$ ,  $p>.05$ ). SSTQUAL 요인은 고객만족( $\beta=.826$ ,  $p<.01$ ), 행동의도( $\beta=.271$ ,  $p<.01$ )에 긍정적인 영향을 미치는 것으로 나타났다. 또한 고객만족은 행동의도에 정(+의 영향을 미치는 것으로 나타났다( $\beta=.521$ ,  $p<.01$ ).

본 연구에서는 기술불안과 행동의도 간의 영향관계에서 SSTQUAL과 고객만족의 다중매개효과를 검증하기 위해 팬텀변수(phantom variable)를 이용하고 부트스트래핑(Bootstrapping) 방법을 활용하여 매개효과의 유의성 검증을 실시하였다. 분석결과, 호텔·외식 SST 이용고객의 기술불안과 행동의도 간의 영향관계에서 SSTQUAL 요인은 매개효과가 있는 것으로 나타났다( $\beta_A$ : 추정치  $=-.105(-.389 \times .271)$ ,  $p<.05$ ). 반면, 고객만족은 매개효과가 없는 것으로 나타났다( $\beta_B$ : 추정치  $=-.028(-.053 \times .521)$ ,  $p>.05$ ).

Table 1. Results of hypotheses tests

Hypotheses	b	$\beta$	S.E.	t	Result
H1 Technology anxiety $\rightarrow$ SSTQUAL	-.239	-.389	.044	-5.427**	Supported
H2 Technology anxiety $\rightarrow$ Satisfaction	-.042	-.053	.039	-1.063	Not supported
H3 Technology anxiety $\rightarrow$ Behavioral intention	-.115	-.136	.039	-2.905**	Supported
H4 SSTQUAL $\rightarrow$ Satisfaction	1.046	.826	.098	10.676**	Supported
H5 SSTQUAL $\rightarrow$ Behavioral intention	.372	.271	.142	2.619**	Supported
H6 Satisfaction $\rightarrow$ Behavioral intention	.564	.521	.110	5.130**	Supported

$\chi^2=680.773$ ,  $df=313$ ,  $\chi^2/df=2.175$ , RMSEA=.065, NFI=.884, IFI=.934, TLI=.925, CFI=.933  
 \* $p<.05$ , \*\* $p<.01$



a: Solid lines indicate significant paths( $\beta$ ), while dotted line indicates non-significant path  
 \* $p<.05$ , \*\* $p<.01$

Figure 2. Paths of multiple mediating effect

#### 4. 결론 및 제언

본 연구는 호텔·외식기업 SST 이용고객의 기술불안과 행동의도 간의 영향관계에서 SSTQUAL과 고객만족의 다중매개효과를 검증하고자 하였다. 고객의 기술불안감과 SST 이용에 대한 행동의도 간의 부정적인 영향관계에서 SSTQUAL은 부분매개효과를 나타내고 있으며, 고객만족은 매개효과가 없는 것으로 나타났다. SST 도입을 통해 기술기반 서비스 인카운터의 변화를 피하고 있는 기업은 고객의 적극적인 행동의도를 높이기 위해서 고객이 지각하는 기술불안감을 감소시킬 수 있는 프로그램이 필요하며, 기능, 즐거움, 고객화, 디자인 등의 셀프서비스기술 서비스품질(SSTQUAL)을 향상시키는 노력을 중요시해야 할 것이다.

#### References

Dabholkar, P. (1996). Consumer evaluations of new technology-based self-service options: an investigation of alternative models of service quality. *International Journal of Research in Marketing*, 13(1), 29-51.

Fornell, C., Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.

Gelbrich, K., Sattler, B. (2014). Anxiety, crowding, and time pressure in public self-service technology acceptance. *Journal of Service Marketing*, 28(1), 82-94.

Ghosh, M.(2020). Understanding the relationship of self-service technology quality with user adoption using SSTQUAL. *The TQM Journal*, 1754-2731. <https://doi.org/10.1108/TQM-12-2019-0291>

- Hair, J., Black, W. C., Babin, B. J., Anderson, R. E. (2010). *Multivariate Data Analysis: A Global Perspective*, Upper Saddle River, New Jersey: Pearson.
- Lin, J. S. C., Hsieh, P. L. (2011). Assessing the self-service technology encounters: Development and validation of SSTQUAL scale. *Journal of Retailing*, 87, 194-206.
- Lyu, F., Lim, H.-A., Choi, J. (2019). Customer acceptance of self-service technologies in retail: A case of convenience stores in China. *Asia Pacific Journal of Information Systems*, 29(3), 428-447.
- Nunnally, J. C. (1978). *Psychometric Theory*(2nd ed.), New York: McGraw-Hill.
- Parasuraman, A., Zeithaml, V. A., Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-37.
- Ribbink, D., Riel, A., Liljander, V., Streukens, S. (2004). Comfort your online customer: quality, trust and loyalty on the Internet. *Managing Service Quality*, 14(6), 446 - 456.
- Safaeimanesh, F., Kılıç, H., Alipour, H., Safaeimanesh, S. (2021). Self-service technologies (SSTs)-the next frontier in service excellence: Implications for tourism industry. *Sustainability*, 13(5), 2604.
- Shin, H., Perdue, R. R. (2019). Self-service technology research: a bibliometric co-citation visualization analysis. *International Journal of Hospitality Management*, 80(3), 101-112, <https://doi.org/10.1016/j.ijhm.2019.01.012>.

## 투자 지표로서 거래량의 유의성에 관한 연구

김영준<sup>1</sup>, 임상빈<sup>2</sup>, 양기성<sup>3</sup>

### 요약

본 연구는 주식 시장에서 거래량(거래 주수 및 거래대금) 지표가 투자전략 관점에서 갖는 정보의 유의성을 분석하는 것을 목적으로 한다. 이에 본 연구는 선행 연구에서 사용된 지표를 포함한 총 7개의 거래량 지표를 국내 주식시장에 적용하여, 투자자 유형(외국인, 기관, 개인)별 순매수가 강한 종목을 매수하고 순매수가 약한 종목을 매도하는 롱-숏 포트폴리오를 구성하여 그 성과를 분석한다. 장기(6개월-1년) 또는 단기(1일) 거래량을 이용한 대부분의 선행연구와는 달리, 본 연구는 포트폴리오 구성 기간을 단기(4주-12주)로 하여 단기 거래량 정보의 유의성을 조사하였으며 분석의 강건성을 위해 구성기간을 단기(5일)로 하여 추가 분석을 진행하였다. 주요 분석 결과는 다음과 같다. 첫째, 분석에 사용된 모든 거래량 지표에 대하여, 롱-숏 포트폴리오의 성과는 개인, 기관, 외국인 순으로 우수하였다. 둘째, 개인 투자자의 순매수 강도를 추종하는 전략은 포트폴리오 구성 기간과 보유 기간에 관계없이 통계적으로 유의한 양의 수익률을 보이고, 대체로 통계적으로 유의한 초과수익률을 보였다. 셋째, 외국인 투자자의 순매수 강도를 추종하는 전략은 대체로 유의한 음의 수익률과 초과 수익률이 보였다. 넷째, 기관 투자자의 순매수 강도를 추종하는 전략은 대체로 통계적으로 유의하지 않았다. 다섯째, 해당 결과를 투자자별 거래행태와 국내시장의 역모멘텀 현상으로 설명하였다. 따라서 본 연구는 장기 및 단기 거래량의 투자지표로서의 유의성을 살펴본 선행연구들이 각기 다른 거래량 척도를 사용하여 일관되지 않은 결과를 보인 것과는 달리, 단기 거래량 정보를 사용하는 경우 거래량 척도의 정의와 관계없이 투자전략 관점에서 일관된 함의를 찾을 수 있는 실증결과를 도출하고 그 결과를 투자자별 거래행태와 국내시장의 역모멘텀 현상을 근거로 해석했다는 점에서 의의가 있다.

<sup>1</sup>06978 서울특별시 동작구 상도로 369, 숭실대학교 금융학부 학사과정. E-mail: 0\_jun@naver.com

<sup>2</sup>06978 서울특별시 동작구 상도로 369, E-mail: ysb9087@soongsil.ac.kr

<sup>3</sup>06978, (교신저자) 서울특별시 동작구 상도로 369, 숭실대학교 금융학부 조교수. E-mail: ksyang@ssu.ac.kr

## Multilevel topic model: a nonparametric multilevel latent class analysis for topic dependency

*Youngsun Kim<sup>1</sup>, Hwan Chung<sup>2</sup>, Saebom Jeon<sup>3</sup>*

### Abstract

A topic model is a probabilistic generative model which discovers latent topics in a collection of document. Latent topics are defined by a multinomial distribution consisting of the probability of the appearance of words and each documents are represented by the proportion of those topics. The topic proportion is a fixed parameter in probabilistic latent semantic analysis (Papadimitriou et al, 1998), whereas it is a random variable from Dirichlet distribution in latent Dirichlet allocation (Blei, 2003). However, it is more reasonable to assume that the topic proportion of each document is governed by the document's overall theme (i.e., document-level latent topic) rather than randomly generated by a probability distribution. Thus, in this study we propose multilevel topic model (MTM), where the word-level topic proportion of each document is determined by the document-level latent topics using multilevel latent class analysis (MLCA) approach. The parameter estimation for MTM has been conducted via modified expectation-maximization (EM) algorithm, where the upward-downward algorithm is applied to alleviate the complexity of calculating the posterior probabilities. Empirical studies on corpus have been conducted on MTM and compared to LDA. These studies have demonstrated that MTM outperforms LDA in terms of model performance and systematic interpretability.

*Keywords* : topic model, latent variable model, multilevel structure.

---

<sup>1</sup>Department of Statistics, Korea University, Seoul, 02841, Korea. E-mail: kim0sun@korea.ac.kr

<sup>2</sup>Department of Statistics, Korea University, Seoul, 02841, Korea. E-mail: hwanch@korea.ac.kr

<sup>3</sup>(Corresponding author) Department of Marketing Bigdata, Mokwon University, Daejeon, 35349, Korea.  
E-mail: alwaysns@mokwon.ac.kr

## Robust Support Vector Machine for Multicategory p-variate functional data

*Jeongeun Sim*<sup>1</sup>, *Seung Jun Shin*<sup>2</sup>

### Abstract

In this article, we propose a robust support vector machine with p-variate functional predictor. The proposed method employs the truncated loss that is known to satisfy the Fisher consistency in multiclass problems. To handle function predictors, we exploit a linear functional kernel defined by the tensor product of each gram matrix of the basis of the functional predictor. A numerical study under various scenarios demonstrates the promising performance of the proposed method.

*Keywords* : Functional data, Multicategory Classification, Truncated Hinge Loss.

---

<sup>1</sup>Graduate Student, Department of Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: sim0348@korea.ac.kr

<sup>2</sup>(Corresponding Author) Associate Professor, Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: sjshin@korea.ac.kr

## On weighted least squares regression with partially interval-censored data

*Yeji Kim*<sup>1</sup>, *Sangbum Choi*<sup>2</sup>

### Abstract

The weighted least squares (WLS) regression method has been widely used to analyze censored survival data in research studies. In this study, our main objective is to develop a new WLS estimation technique with inverse probability of censoring weighting (IPCW) to estimate coefficients in a linear regression model. Unlike classical survival analysis methods like the Cox and accelerated failure time (AFT) models, the linear regression model considers the effect of covariates and provides easy-to-understand interpretations. It is also simple to compute and can produce stable estimates in practical scenarios. Our research focuses on doubly-censored (DC) data and partly interval-censored (PIC) data. We propose a weighted linear regression model that incorporates weights for each data type, and we estimate the variance using theoretical methods that demonstrate sufficient asymptotic properties. Moreover, we provide a method for selecting important regression coefficients using LASSO, ALASSO, and SCAD methods. We also perform extensive simulations to evaluate the performance of our proposed estimators in finite samples and apply our approach to data from Nigerian children's mortality data.

*Keywords* : Accelerated lifetime, Partly interval censoring, Semiparametric model, Inverse probability weighting, linear regression model, Survival analysis.

---

<sup>1</sup>Statistics, Korea University, 145 Anam-ro, Seunbuk-gu, Seoul 02841, South Korea.

E-mail: yeji\_kim@korea.ac.kr

<sup>2</sup>(Corresponding Author) Statistics, Korea University, 145 Anam-ro, Seunbuk-gu, Seoul 02841, South Korea. E-mail: choisang@korea.ac.kr

## 모바일 게임에서 이탈 조짐 유저에 대한 조기 탐지

이민혁<sup>1</sup>, 김수인<sup>2</sup>, 박선우<sup>3</sup>, 이성환<sup>4</sup>, 정윤서<sup>5</sup>

### 요 약

유저 이탈 예측은 게임사에게 중요한 과제 중 하나이다. 모바일 게임 산업에서 유저 잔존율은 서비스 수명 및 수익과 연관이 깊기 때문이다. 만약, 특정 유저의 이탈을 사전에 예측할 수 있다면 프로모션을 통해 해당 유저의 잔존을 유도할 수 있을 것이다. 따라서, 이탈 예측 문제에서는 예측의 정확도도 중요하지만 이탈의 징후를 얼마나 빠르게 파악할 수 있는지 또한 중요하다. 본 연구에서는 이탈 징후를 조기에 탐지하기 위하여, 유저별 잔존 확률을 일별로 예측하고 이 예측된 확률값들을 활용하여 유저 이탈 징후를 조기에 파악하는 방법을 제안하고자 한다. 이를 위해, 카카오 게임즈 사의 오딘 유저 로그 데이터로 Long Short Term Memory(LSTM) 모델을 학습하고 유저별 잔존 확률을 구하였다. 잔존 확률의 변화 패턴에 대한 분석을 통해 이탈 가능성이 높은 유저를 조기에 감지할 수 있는 경험적 규칙을 제시하였다. 성능 평가 결과, 기존에 접속일을 기반으로 한 규칙보다 제시한 규칙을 이용할 시 이탈 유저를 조기에 감지할 수 있음을 확인하였다.

주요용어 : 모바일 게임, 순환 신경망, 이탈 예측, LSTM.

---

<sup>1</sup>02841, 서울특별시 성북구 안암로 145 고려대학교 정경대학 통계학과 석사과정.  
E-mail: leeminhyuk12@korea.ac.kr

<sup>2</sup>02841, 서울특별시 성북구 안암로 145 고려대학교 정경대학 통계학과 석박통합과정.  
E-mail: ksi2002@korea.ac.kr

<sup>3</sup>02841, 서울특별시 성북구 안암로 145 고려대학교 정경대학 통계학과 석사과정.  
E-mail: 2022020386@korea.ac.kr

<sup>4</sup>02841, 서울특별시 성북구 안암로 145 고려대학교 정경대학 통계학과 석사과정. E-mail: mmbv83@korea.ac.kr

<sup>5</sup>(교신저자) 02841, 서울특별시 성북구 안암로 145 고려대학교 정경대학 통계학과 교수.  
E-mail: yoons77@korea.ac.kr



## Proformer: a hybrid macaron transformer model predicts expression values from promoter sequences<sup>\*</sup>

*Il-youp Kwak*<sup>1</sup>, *Byeong-Chan Kim*<sup>1</sup>, *Juhyun Lee*<sup>1</sup>, *Daniel J Garry*<sup>2</sup>, *Wuming Gong*<sup>2</sup>

### Abstract

The breakthrough high-throughput measurement of the cis-regulatory activity of millions of randomly generated promoters provides an unprecedented opportunity to systematically decode the cis-regulatory logic that determines the expression values. We developed an end-to-end transformer encoder architecture named Proformer to predict the expression values from DNA sequences. Proformer used a Macaron-like Transformer encoder architecture, where two half-step feed forward (FFN) layers were placed at the beginning and the end of each encoder block, and a separable 1D convolution layer was inserted after the first FFN layer and in front of the multi-head attention layer. The sliding k-mers from one-hot encoded sequences were mapped onto a continuous embedding, combined with the learned positional embedding and strand embedding (forward strand vs. reverse complemented strand) as the sequence input. Moreover, Proformer introduced multiple expression heads with mask filling to prevent the transformer models from collapsing when training on relatively small amount of data. We empirically determined that this design had significantly better performance than the conventional design such as using the global pooling layer as the output layer for the regression task. These analyses support the notion that Proformer provides a novel method of learning and enhances our understanding of how cis-regulatory sequences determine the expression values.

*Keywords* : Transformer, Gene expression data, deep learning.

---

<sup>\*</sup>This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science (RS-2023-00208284).

<sup>1</sup>(Corresponding author) 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Department of Applied Statistics, Chung-Ang University. E-mail: ikwak2@cau.ac.kr

<sup>2</sup>420 Delaware Street SE, Minneapolis, MN 55455, Cardiovascular Division, Department of Medicine, University of Minnesota.

## 클래스 불균형 데이터 분류를 위한 군집화 기반 언더샘플링 연구

김주미<sup>1</sup>, 정여진<sup>2</sup>

### 요약

최근 데이터 저장 기법의 비약적인 발달과 처리비용의 감소에 힘입어 머신러닝에 기반을 둔 분류예측 문제를 많이 시도하고 있다. 그러나 신용카드 사기 탐지나 고객이탈, 장애탐지 등 일반적이지 않은 이상탐지 분야에서는 다수클래스와 소수클래스가 불균형하게 분포하며 분류 성능에 많은 오류를 야기한다. 이를 해결하기 위한 데이터 조정 접근 방식 중 Under/Over 샘플링에 관한 연구가 활발히 진행되어 왔으며 많은 성과를 이루었다. 본 논문에서는 빅데이터 처리를 위한 현실적인 Unders 샘플링 방안을 고민했으며 시뮬레이션을 통해 Gaussian Distribution을 따르는 불균형 자료를 생성하고, 다수클래스의 모집단 분포를 반영하는 cluster 기반 샘플링 방식을 진행했다. 또한 Kullback-Leibler divergence을 활용하여 도출된 subset이 모집단과 확률분포가 같은 집단임을 검증하는 시도를 통해 Under 샘플링의 최대 단점인 정보손실이 최소화된 subset임을 확인하였다. 성능평가를 위해 XGB모형을 통해 각 샘플링 기법별 분류예측을 시도한 결과, 제안하는 기법이 기존 기법에 비해 성능이 우수함을 확인할 수 있었는데 특히 데이터의 불균형의 정도가 클수록 그 차이가 크게 나타났다.

주요용어 : 불균형데이터(Imbalanced Data), Undersampling, 군집(Cluster), 쿨백라이블러발산(Kullback-Leibler divergence), XGBoost (eXtreme Gradient Boosting)

---

<sup>1</sup>02707 서울특별시 성북구 정릉로77 국민대학교 데이터사이언스학과 박사과정. E-mail: jumi.kim@kookmin.ac.kr  
<sup>2</sup>(교신저자) 02707 서울특별시 성북구 정릉로77 국민대학교 데이터사이언스학과 교수.  
E-mail: ychung@kookmin.ac.kr

## Clustering based Undersampling for Imbalanced Data Classification

*Ju Mi Kim*<sup>1</sup>, *Yeo Jin Chung*<sup>2</sup>

### Abstract

In recent years, due to the rapid development of data storage techniques and the reduction of processing costs, there has been a growing interest in applying machine learning-based approaches to tackle classification problems. However, in the field of anomaly detection that is not common, such as credit card fraud detection, customer attrition and failure detection, the major and minor classes are unevenly distributed and cause substantial errors in classification performance. To address this issue, extensive research has been conducted on data level approach, with particular focus on undersampling and oversampling techniques and many achievements have been made. This study focuses on exploring practical undersampling approaches for big data processing. Through simulations, imbalanced data that follows Gaussian Distribution is generated and a cluster-based sampling method that reflects the distribution of the population is conducted. Furthermore, using Kullback-Leibler Divergence, an attempt was made to validate that the subset derived from the proposed method shares the same probability distribution as the population. This validation process aimed to construct a subset with minimal information loss, effectively addressing the major drawback of undersampling. Finally, the classification prediction performance of each sampling method was evaluated using the XGB model and the results confirmed that the proposed method outperformed existing approaches, particularly when the degree of imbalance in the data is large.

*Keywords* : Imbalanced Data, Undersampling, Cluster, XGBoost, Kullback-Leibler divergence.

---

<sup>1</sup>Ph. D.student Department of Data Science, Kookmin University, Jeongneung-ro77, Seoul, Korea.

E-mail: jumi.kim@kookmin.ac.kr

<sup>2</sup>(Corresponding Author) Professor, Department of Data Science, Kookmin University, Jeongneung-ro77, Seoul, Korea. E-mail: ychung@kookmin.ac.kr

## Mixture modelling for analyzing a rainfall pattern with Julia: A case study in South Korea\*

*Hee-Young Kim*<sup>1</sup>, *Su-Hyeon Kim*<sup>2</sup>, *Ji-Yeon Oh*<sup>3</sup>

### Abstract

“Heavy precipitation” refers to instances during which the amount of rain or snow experienced in a location substantially exceeds what is normal. What constitutes a period of heavy precipitation varies according to location and season. Climate change can affect the intensity and frequency of precipitation. This study investigates the maximum yearly precipitation amounts reported by stations from 1961 to 2020 for Seoul, Busan, Gwangju, and Ulsan. The data are not only large but also complex. The conventional rainfall analysis has been routinely used univariate Gumbel distribution for rainfall analysis in South Korea. A natural question to ask therefore is how the distribution of extreme rainfall have been observed by the effect of climate change. Therefore, in this article, we analyze the data with several mixture models with Gumbel, Weibull, Normal, Lognormal, and Gamma. The selection of the best model is AIC and BIC. In particular, we implement the method using the Julia language (Bezanson, Edelman, Karpinski, and Shah 2017), which is the most recent new language, and is designed for efficient and parallel numerical computing while keeping a high level of human readability.

*Keywords* : Rainfall amounts; Mixture modeling; Gumbel distribution.

---

\*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).

<sup>1</sup>(Corresponding Author) Professor, Bigdata Science, Korea University, Sejong-Street 2511 Sejong-City 339-700, E-mail : starkim@korea.ac.kr

<sup>2</sup>Undergraduate student, Department of National Statistics, Korea University, Sejong-Street 2511 Sejong-City 339-700.

<sup>3</sup>Undergraduate student, Department of National Statistics, Korea University, Sejong-Street 2511 Sejong-City 339-700.

## Imputation Methods For Multivariate Longitudinal Data Using Mixture Of Generalized Linear Mixed Models

*Daegyeong Yang*<sup>1</sup>, *Juwon Song*<sup>2</sup>

### Abstract

Longitudinal data often collect many variables that may include discrete as well as continuous variables. Many longitudinal data inherently suffer from missing values due to the nature of their collections. However, most imputation methods focus on imputation of univariate data. On the other hand, when longitudinal data are imputed, it is usually assumed that subjects have a common mean trajectory. When there exist several different mean trajectories in the data, Yang and Song (2022) proposed an imputation method for univariate longitudinal data in the presence of trajectory heterogeneity. This study extends their method to handle multivariate longitudinal data in which both continuous and discrete variables are observed longitudinally. Simulation studies are conducted to evaluate performance of the suggested imputation method.

*Keywords* : Multivariate longitudinal data, Missing data, Generalized linear mixed model, Clustering.

---

<sup>1</sup>The Integrated Master & Ph.D Course, Department of Statistics, Korea University, 145 Anam-Ro, Sungbuk-Gu, Seoul 02841, Korea. Email: rudl77@korea.ac.kr

<sup>2</sup>Professor, Department of Statistics, Korea University, 145 Anam-Ro, Sungbuk-Gu, Seoul 02841, Korea. Email: jsong@korea.ac.kr

## Introduction to Privacy Preserving Data Sharing and Analytics with Homomorphic Encryption

*Sulgi Kim*<sup>1</sup>, *Soo-Heang Eo*<sup>2</sup>, *Garam Lee*<sup>3</sup>

### Abstract

As more data is being analyzed and utilized in the real world, unexpected breaches of sensitive data such as privacy data are increasing. For example, more than 50 million records are being leaked every year only in the US healthcare domain. Therefore, various privacy enhancing technologies (P.E.Ts) have attracted attention including homomorphic encryption (HE), secure multi-party computation, differential privacy and trusted execution environment. Accordingly, the White house published recent report regarding privacy preserving data sharing and analytics (PPDSA), which is also referred to as privacy preserving machine learning (PPML). In this talk, we introduce homomorphic encryption and its application in PPDSA. As one of the P.E.Ts, Homomorphic Encryption is a specific encryption technique that allows numerical operations in an encrypted state. Since HE allows whole data processing pipeline without ever decrypting or exposing data, it can obtain competing goals of utility and security. Since 2016, when an advanced HE scheme, CKKS has been introduced, there has been dramatic improvement in efficiency. Now, we have use cases in real world with homomorphic encryption including a COVID-19 tracing service using personal location data. PPDSAs using HE are expected to provide good use cases in military, security agency, government, healthcare and financial fields.

*Keywords* : PPDSA, Private Enhancing Technology, PPML.

---

<sup>1</sup>Director, CryptoLab Inc., 1 Gwanak-ro Gwanak-gu, Seoul, Korea, 08826. E-mail: [sulgik@cryptolab.co.kr](mailto:sulgik@cryptolab.co.kr)

<sup>2</sup>Product Evangelist, CryptoLab Inc., 1 Gwanak-ro Gwanak-gu, Seoul, Korea, 08826.

E-mail: [abel.eo@cryptolab.co.kr](mailto:abel.eo@cryptolab.co.kr)

<sup>3</sup>Evangelist, CryptoLab Inc., 1 Gwanak-ro Gwanak-gu, Seoul, Korea, 08826.

E-mail: [garamlee@cryptolab.co.kr](mailto:garamlee@cryptolab.co.kr)

## Privacy Preserving Machine Learning for Secure Outsourced Computing

*Soo-Heang Eo*<sup>1</sup>, *Garam Lee*<sup>2</sup>, *Sulgi Kim*<sup>3</sup>

### Abstract

The use of machine learning for prognostic and health management (PHM) has shown great promise in improving the efficiency and accuracy of predictive maintenance. Nonetheless small and medium-sized manufacturing enterprises (SMEs) frequently lack the required expertise and resources to conduct in-house PHM analytics. While cloud-based services offer these SMEs the option to outsource PHM analytics, a major obstacle to this approach is the reluctance of data owners to share their data due to concerns about data privacy. To address these concerns, privacy-preserving machine learning techniques have emerged as a promising solution for secure PHM outsourcing. In this presentation, we introduce homomorphic encryption as a promising technique to achieve secure PHM outsourcing with high precision for SMEs. Homomorphic encryption is a cryptographic technique that allows direct computation on encrypted data, enabling the analysis of sensitive data while preserving its privacy (Cheon et al, 2017). We showcase a two-party collaborative framework that utilizes homomorphic encryption for secure PHM outsourcing of analytics for SMEs (Kang et al., 2021). This framework enables data owners to keep their sensitive data private while allowing a third-party service provider to perform PHM analytics. By utilizing homomorphic encryption, the data remains encrypted throughout the computation process, ensuring the protection of the data owner's privacy. In summary, we demonstrate how homomorphic encryption can facilitate secure PHM outsourcing for SMEs, providing a solution that addresses data privacy concerns and data analysis precision. Furthermore, we also provide an overview of the strengths and limitations of the approach used in this study.

Keywords: PPML, Private Enhancing Technology, Secure PHM.

---

<sup>1</sup>Product Evangelist, CryptoLab Inc., 1 Gwanak-ro Gwanak-gu, Seoul, Korea, 08826.  
E-mail: abel.eo@cryptolab.co.kr

<sup>2</sup>Evangelist, CryptoLab Inc., 1 Gwanak-ro Gwanak-gu, Seoul, Korea, 08826.  
E-mail: garamlee@cryptolab.co.kr

<sup>3</sup>Director, CryptoLab Inc., 1 Gwanak-ro Gwanak-gu, Seoul, Korea, 08826. E-mail: sulgik@cryptolab.co.kr

## Application of homomorphic encryption in predicting postoperative mortality using multi-institutional data

*Garam Lee<sup>1</sup>, Jungyo Suh<sup>2</sup>, Sang-Wook Lee<sup>3</sup>, Junbum Shin<sup>4</sup>, Sulgi Kim<sup>5</sup>*

### Abstract

Due to the potential for personal data leakage, medical data sharing among institutions remains severely limited. Homomorphic encryption, a cryptographic primitive enabling computations on encrypted data, provides a viable solution to overcome these limitations, notably in highly regulated domains such as the medical field. This technology paves the way for pioneering new medical services by facilitating secure data sharing. Our study explores the feasibility of merging diverse datasets using homomorphic encryption and training machine learning models on these integrated datasets using real multi-institutional data. We also verify the potential for improved predictive performance through dataset merging, thereby underscoring the utility of homomorphic encryption for multi-institutional datasets. Moreover, we strive to ascertain the optimal volume of internal hospital data necessary to develop efficient in-house prediction models, achieved by progressively supplementing our internal datasets with a substantial external dataset. We utilized data from 341,008 patients aged 18 or older, who underwent non-cardiac surgery across three medical institutions. Our primary outcome measure was in-hospital mortality within 30 days post-surgery. Secure logistic regression served as our predictive modeling method. We compared the predictive performances of individual institutional models with those trained on multi-institutional integrated datasets processed via homomorphic encryption. Our results revealed enhanced performance from models trained on the multi-institutional data fused using homomorphic encryption technology, compared to those based on single-center datasets. Furthermore, augmenting internal hospital datasets with large-scale external data enhanced model efficacy.

*Keywords:* Homomorphic Encryption, Privacy-preserving Machine Learning (PPML) Multi-institutional Data, Secure Logistic Regression, Postoperative Mortality.

---

<sup>1</sup>CryptoLab Inc., Seoul

<sup>2</sup>Asan Medical Center, University of Ulsan College of Medicine, Seoul

<sup>3</sup>Asan Medical Center, University of Ulsan College of Medicine, Seoul

<sup>4</sup>CryptoLab Inc., Seoul

<sup>5</sup>CryptoLab Inc., Seoul



## Land use regression method as sign constrained regularized regressions

*Soon-Sun Kwone*<sup>1</sup>, *Hosik Choie*<sup>2</sup>, *Whanhee Lee*<sup>3</sup>,  
*Yeonjin Kime*<sup>4</sup>, *Hwan-Cheol Kime*<sup>5</sup>, *Woojoo Lee*<sup>6</sup>

### Abstract

Land-use regression is a popular method for predicting ambient pollutant concentrations at points of interest where no measurements are taken. LUR has been used to interpolate ambient pollutant concentrations at points of interest where no measurements are available using measurements from several monitoring sites. For example, it was used to predict outdoor concentrations at the home addresses of the study participants in more than 30 cohort studies (Beelen, 2011). The small-area variation in air quality and health (SAVIAH) study was the first to use LUR to model small-scale variations in air pollution (Briggs, 1997). It included variables, such as the traffic volume, population density, land use, and altitude, which were compiled using geographic information systems (GIS). The LUR model was constrained by requiring that all the regression coefficients have the a priori definition of the required signs for regression coefficients. Due to this point, the model-building process is complicated, and systematically understanding when and how the process works is difficult. To overcome these limitations, we reformulate the existing land use regression method as a sign-constrained regression problem with an explicit objective function to be minimized. This novel formulation always leads to estimated regression coefficients that satisfy the predefined direction based on subject matter knowledge while simultaneously substantially improving the prediction performance of the existing land-use regression method. The advantages of the proposed sign-constrained regression method are confirmed through a numerical study and real data analysis.

Keywords: Land use regression, Sign constraints, Penalized regression methods, Prediction, Interpretability.

---

<sup>1</sup>First author, Department of Mathematics/Artificial Intelligence, Ajou University, 206 World cup-ro, Suwon, 16499, Gyeonggi, South Korea. E-mail: qrio1010@ajou.ac.kr

<sup>2</sup>Department of Urban Big Data Convergence, University of Seoul, 163 Seoulsiripdae-ro, Seoul, 02504, South Korea

<sup>3</sup>School of Biomedical Convergence Engineering, Pusan National University, 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, 46241, South Korea.

<sup>4</sup>Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea.

<sup>5</sup>Department of Occupational and Environmental Medicine, Inha University, School of Medicine, 100 Inha-ro, Michuhol-gu, Incheon 22212, South Korea.

<sup>6</sup>Correspondence, Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea. E-mail: lwj221@snu.ac.kr

## 간호대학생의 비판적 사고성향과 간호정보역량이 간호과정 수행능력에 미치는 영향

김민영<sup>1</sup>

### 요 약

본 연구는 간호대학생의 비판적 사고성향, 간호정보역량, 간호과정 수행능력의 정도를 확인하고 이들의 상관관계를 파악하기 위하여 시행되었다. 본 연구는 부산, 경남에 소재한 5개 대학의 간호학과 2, 3, 4학년생을 대상으로 자료수집에 대한 동의한 학생을 대상으로 자료를 수집하였고, 수집된 자료는 SPSS program 23.0을 이용하여 분석하였다. 연구결과 비판적 사고성향은 평균  $3.61 \pm 0.45$ 점, 간호정보역량은 평균  $3.56 \pm 0.54$ 점, 간호과정 수행능력은 평균  $3.68 \pm 0.57$ 점으로 나타났다. 연구 대상자의 일반적 특성에 따른 비판적 사고성향은 성별, 전공만족도에서 유의한 차이를 보였으며, 간호정보역량은 형제 수에 따라 통계적으로 유의한 차이를 보였고, 간호과정 수행능력은 학년, 전공 만족도에서 통계적으로 유의한 차이를 보였다. 상관분석에서 비판적 사고성향과 간호정보역량 ( $r=.324, p<.001$ ), 간호과정 수행능력과 비판적 사고성향( $r=.600, p<.001$ ), 간호정보역량과 간호과정 수행능력( $r=.570, p<.001$ )은 통계적으로 유의미한 정적 상관관계가 있는 것으로 나타났다. 대상자의 간호과정 수행능력에 미치는 영향요인을 알아보기 위해 비판적 사고성향과 간호정보역량 및 일반적 특성에서 간호과정 수행능력에 유의한 차이를 나타낸 학년과 전공만족도를 분석한 결과 비판적 사고성향( $\beta=.407, p<.001$ ), 전공만족도( $\beta=.139, p=.011$ ), 학년( $\beta=.160, p=.003$ )이 영향요인으로 나타났다.

주요용어 : 간호대학생, 비판적 사고성향, 간호정보역량, 간호과정 수행능력.

### 1. 연구의 필요성

최근 보건의료 환경은 급속히 진행되는 고령화와 저출산, 건강 불평등 심화, 자연적 및 사회적 재난 발생 등으로 인해 역동적으로 변화되고 있다(Seo, Hong, 2022; Kim, Lee, 2018). 이러한 시대적 요구에 따라 한국간호교육평가원에서는 국내 의 의료 현장에서 요구되는 간호사 역량을 갖춘 학생을 배출할 수 있도록 지속적인 간호교육 프로그램 개선과 간호교육의 질 향상을 권장하고 있다(Seo, Hong, 2022; Park, 2015).

현재 대학에서의 간호학 교육은 보건의료 환경과 정보화 시대에 맞추어 학생의 창의력을 향상시키고 사고력을 길러주며 새로운 지식을 창출하는 전문적인 직업인의 필수 능력으로서 비판적 사고 성향을 강조하고 있다(Park, Cho, Park, 2015; Yune, Shin, Lee, 2007). 간호사는 복잡하고 빠르게 변화하는 간호현장에서 비판적 사고성향을 가지고 능력을 발휘할 때 의사결정과 신중한 대처, 자율성을 유지 할 수 있다(Kim, lee, 2016; Shin, Hwang, Shin, 2008). 이에 많은 간호교육자들은 임상

<sup>1</sup>47011 부산광역시 사상구 주례로 45, 경남정보대학교 간호학과 조교수. E-mail: 0502young@hanmail.net

상황에서 학생들의 비판적 사고력 개발을 촉진 시킬 수 있는 학습 설계와 전략 개발에 관심을 두고 있다(Seo, Hong, 2022; Kwon, Seo, 2012).

또한 21세기 지식기반 사회에서는 간호사가 필요한 정보를 탐색, 수집, 평가, 분석하여 종합할 수 있는 능력인 정보활용능력을 갖추는 것이 강조되고 있다(Jang, Kim, Kim, 2015). 간호정보역량이란 간호업무에서 정보와 관련된 역할을 수행하는 능력을 일컫는다(Kwon et al., 2016; Kim et al., 2013). 이는 컴퓨터 프로그램을 활용하는 능력뿐만 아니라 정보화와 관련된 지식과 자료를 활용하고 간호에 적용하는 것이다(Kwon et al., 2016; Staggers et al., 2002). 간호사는 환자상태에 대한 정확한 판단과 환자에게 발생할 수 있는 문제를 예측하고 대처할 능력이 있는 전문성이 높은 임상수행능력을 갖추어야 한다(Choi et al., 2013; Jang, 2000).

간호과정은 간호학적 지식을 기반으로 대상자의 문제를 파악하여 그 문제를 해결해주기 위한 필수적이고 과학적인 간호사의 수행방식인 간호전문 활동의 영역을 수행하는 과정으로(Hwang, Jeon, 2019; Yang, Sim, 2018) 간호 대학생은 이론적 지식을 바탕으로 임상실습을 통해 전문직 간호사로 성장하기 위해 필요한 다양한 지식, 기술, 행위를 실무에 적용할 수 있도록 배우게 된다(Hwang, Jeon, 2019; Yang, Park, 2004). 간호사에게 요구되는 지식과 기술의 범주도 다양해지고 있어 환자상태에 대한 정확한 판단과 환자에게 발생할 수 있는 문제를 예측하고 대처할 능력이 있는 전문성이 높은 임상수행능력이 필요하다(Choi et al., 2013; Jang, 2000). 따라서 졸업 전에 간호대학생들은 대상자의 건강문제에 통합적으로 접근할 수 있는 숙련된 임상수행능력을 갖추어야 한다(Choi et al., 2013).

이에 따라 본 연구는 간호대학생의 비판적 사고성향, 간호정보역량, 간호과정 수행능력의 정도에 대해 알아보고, 간호과정 수행능력에 영향을 미치는 요인을 파악하여 간호대학생의 간호과정 수행능력을 발전시키기 위한 전략을 개발하는 데 필요한 기초자료로 활용하고자 시도되었다.

## 2. 연구방법

### 2.1. 연구대상 및 자료수집

본 연구의 대상자는 부산, 경남에 소재한 5개 대학의 간호학과에 재학 중인 1학년을 제외한 간호대학생 2, 3, 4학년을 대상으로 연구의 목적을 이해하고 설문지 작성에 동의한 학생들을 대상으로 설문에 응답하도록 하였다. Park(2018)에 의하면 시뮬레이션 실습 후 임상수행능력, 비판적 사고성향이 통계적으로 유의한 향상을 나타내 1학년은 실습을 경험하지 못한 점과 비판적 사고와 관련된 과목을 수강하지 않은 점에서 비판적 사고 성향과 간호과정 수행능력에 대한 자기 평가가 어렵다고 판단되어 대상에서 제외하였다. 설문지 탈락률을 고려하여 168명을 대상으로 하였다.

### 2.2. 연구도구

#### 2.2.1. 비판적 사고성향

윤진(2004)이 개발한 비판적 사고성향 측정도구를 사용하였다. 각 문항은 5점 Likert 척도로 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’ 5점까지로 되어 있어 점수가 높을수록 비판적 사고성향이 강한 것을 의미한다. 개발 당시 도구의 신뢰도는 Cronbach's  $\alpha$  0.84이었고, 본 연구에서는 Cronbach's  $\alpha$  값은 0.83이었다.

### 2.2.2. 간호정보역량

Stagger(1994)가 간호사 대상의 정보화 수준을 측정하기 위하여 개발한 도구를 2019년 최은주 등(2019)이 간호대학생 대상으로 재개발한 도구를 사용하였다. 총 3개 영역의 25문항으로 5점 likert 척도를 사용하여 ‘전혀 그렇지 않다.’ 1점, ‘그렇지 않다.’ 2점, ‘보통이다.’ 3점, ‘대체로 그렇다.’ 4점, ‘매우 그렇다.’ 5점으로 총 점수가 높을수록 간호정보역량 수준이 높은 것을 의미한다. Stagger(1994)의 개발 당시 전체 Cronbach’s  $\alpha$ 는 .90, 최은주 등(2019)의 연구에서 Cronbach’s  $\alpha$ 는 .90 이었고, 본 연구에서 Cronbach’s  $\alpha$ 는 .87이었다.

### 2.2.3. 간호과정 수행능력

김경윤과 이은주(2016)가 개발한 도구를 사용하였으며 총 17문항으로 구성되어 있다. 점수는 Likert 5점 척도로 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’까지의 5점으로, 점수가 높을수록 간호과정 수행에 대한 능력이 높은 것을 의미한다. 도구의 신뢰도는 개발당시 Cronbach’s  $\alpha$ 는 .94, 본 연구에서 Cronbach’s  $\alpha$ 는 .82이었다.

## 2.3. 자료분석방법

수집된 자료는 SPSS program 23.0을 이용하여 분석하였다. 대상자의 일반적 특성을 실수와 백분율로 구하였고, 비판적 사고성향, 간호정보역량, 간호과정 수행능력은 평균, 표준편차, 최솟값, 최댓값을 산출하였다. 일반적 특성에 따른 비판적 사고성향, 간호정보역량, 간호과정 수행능력은 t-test와 ANOVA를 이용하여 분석하고 사후 분석을 위해 Scheffe’s test을 사용하였다. 비판적 사고성향, 간호정보역량, 간호과정 수행능력의 상관관계를 파악하기 위해 Pearson’s correlation coefficient로 분석하였다.

## 3. 연구결과

### 3.1. 대상자의 일반적 특성

대상자 총 168명 중 연령은 19세부터 25세 이하가 110명(70.2%)으로 가장 많았고, 성별은 여성이 148명(88.1%)으로 대다수를 차지하였다. 학년은 3학년 96명(57.1%)으로 가장 많았고, 거주유형은 부모님과 거주가 114명(67.9%)으로 가장 많았고, 형제자매 유형은 형제자매 수가 1명이 77명(45.8%)으로 가장 많다. 평균학점은 3.1~3.9이하가 109명(64%)으로 가장 많았고, 전공 만족도는 보통이 74명(44%)로 가장 많았다.

### 3.2. 대상자의 비판적 사고성향, 간호정보역량 및 간호과정 수행능력

대상자의 비판적 사고성향의 전체평균은 5점 만점에  $3.61 \pm 0.45$ 점으로 나타났다. 간호정보역량은 5점 만점에  $3.56 \pm 0.54$ 점으로 나타났으며, 간호과정 수행능력은 5점 만점에  $3.68 \pm 0.57$ 점으로 나타났다.

### 3.3. 대상자의 일반적 특성에 따른 비판적 사고성향, 간호정보역량 및 간호과정 수행능력

대상자의 일반적 특성에 따른 비판적 사고성향은 성별( $F=-1.704, p<.05$ )과 전공 만족도( $F=4.668, p<.001$ )에 따라 통계적으로 유의한 차이가 있는 것으로 나타났으며, 사후분석결과 전공 만족도는 보통( $3.52\pm 0.44$ ), 만족( $3.61\pm 0.43$ )에 비해 매우 만족( $3.99\pm 0.40$ )이 유의하게 높게 나타났다. 대상자의 일반적 특성에 따른 간호정보역량은 형제 수에 따라 통계적으로 유의한 차이가 있는 것으로 나타났으며( $F=2.640, p<.05$ ), 형제 수가 3명 이상( $3.32\pm 0.55$ )일 때보다 1명일 때가( $3.64\pm 0.54$ ) 유의하게 높았다. 대상자의 일반적 특성에 따른 간호과정 수행능력은 학년( $F=7.867, p<.001$ )과 전공 만족도( $F=3.789, p<.01$ )에 따라 통계적으로 유의한 차이가 있는 것으로 나타났으며, 학년의 경우 2학년( $3.36\pm 0.66$ )보다는 3학년( $3.74\pm 0.51$ )과 4학년( $3.84\pm 0.52$ )에서 유의하게 높게 나타났고, 전공 만족도의 경우 보통( $3.54\pm 0.51$ )에 비해서 매우 만족( $4.05\pm 0.55$ )한 경우에 유의하게 높았다.

### 3.4. 대상자의 비판적 사고성향, 간호정보역량 및 간호과정 수행능력의 상관관계

대상자의 비판적 사고성향, 간호정보역량 및 간호과정 수행능력과의 상관관계 분석 결과, 비판적 사고성향과 간호정보역량 ( $r=.324, p<.001$ ), 간호과정 수행능력과 비판적 사고성향( $r=.600, p<.001$ ), 간호정보역량과 간호과정 수행능력( $r=.570, p<.001$ )은 통계적으로 유의미한 정적 상관관계가 있는 것으로 나타났다.

### 3.5. 연구대상자의 간호과정 수행능력에 미치는 영향요인

대상자의 간호과정 수행능력에 미치는 영향요인을 알아보기 위해 비판적 사고성향과 간호정보역량 및 일반적 특성에서 간호과정 수행능력에 유의한 차이를 나타낸 학년과 전공만족도를 분석한 결과 비판적 사고성향( $\beta=.407, p<.001$ ), 전공만족도( $\beta=.139, p=.011$ ), 학년( $\beta=.160, p=.003$ )이 영향요인으로 나타났다. 이들 변인에 의한 설명력은 27.9%( $F=22.54, p<.001, R^2=.292, \text{Adj-}R^2=.279$ )이었다. 분석 전 잔차들 간 상관성이 있는지를 알아보기 위해 더빈-왓슨 통계량을 살펴본 결과 1.809로 잔차들 간 상관은 없었으며, 독립변수들 간 다중공선성을 알아보기 위해 공차한계와 분산팽창지수(VIF)를 살펴본 결과 공차한계는 .894~.963으로 .10 이상으로 나타났고, 분산팽창지수(VIF)는 1.036~1.119로 10 이하의 결과를 보여 독립변수들 간 다중공선성은 없었다.

## 다문화청소년의 문화적응스트레스가 학교생활에 미치는 영향: 친구의지지와 자아존중감의 이중매개효과

박공주<sup>2</sup>

### 요 약

본 연구는 다문화청소년의 문화적응스트레스가 학교생활에 미치는 영향을 분석하고 친구의지지와 자아존중감의 매개효과를 검증하고자 한다. 이를 위해 2019 다문화청소년패널조사 2기 자료(MAPS 2019)를 활용하였고, 초등학교 4학년 총 2,271명의 자료를 분석하였다. 자료분석은 SPSS 26.0으로 기술통계분석과 상관관계를 실시했으며, Process Macro 4.0을 활용하여 매개효과 분석을 실시하였다. 다문화청소년의 문화적응스트레스가 학교생활적응의 관계에서 친구의지지와 자아존중감의 매개효과를 검증한 결과 문화적응스트레스가 학교생활적응에 대해 유의미한 영향을 미쳤고, 문화적응스트레스가 학교생활적응에 미치는 영향에서 친구의지지를 매개하여 검증한 결과 유의미한 정적 매개효과를 미치는 것으로 나타났고 자아존중감의 관계에 있어서도 유의미한 매개효과를 미치는 것으로 나타났다. 이를 바탕으로 문화적응스트레스가 학교생활적응에 미치는 영향을 파악하기 위한 다양한 교육 및 상담프로그램 개발에 이바지해야 할 것이다.

### 1. 연구모형과 가설

본 연구는 2019년 한국청소년정책연구원 다문화청소년패널조사 2기 자료(MAPS 2019)의 원시자료를 이용하여 청소년 다문화청소년의 문화적응스트레스가 친구의지지와 자아존중감을 매개하여 학교생활적응에 영향을 미치는지 살펴 보고자 한다. 연구목적은 다음과 같다 <그림 1>.

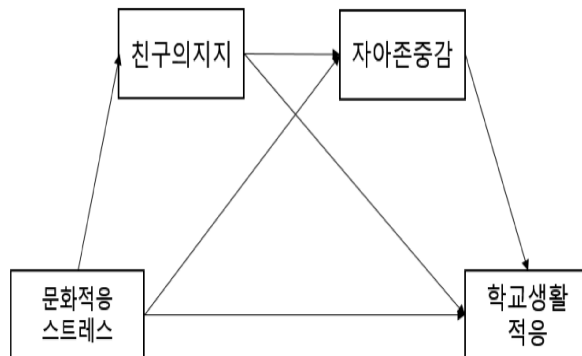


Figure 1. Research model

## 2. 분석자료 및 연구대상

본 연구는 한국청소년정책연구원 2019 다문화청소년패널조사 2기 자료(MAPS 2019)를 활용하였다. MAPS 2기는 다문화청소년의 성장과 발달과정에 대한 기초자료를 구축하는 목적에서 2019년을 기준으로 초등학교 4학년이다. 한국아동·청소년패널조사의 연구대상자 표집방법은 확률비례추출법으로 추출하였다.

## 3. 측정도구

### 1) 독립변수 : 문화적응스트레스

문화적응스트레스는 한국청소년정책연구원 다문화청소년패널조사의 (Hovey & King(1996)의 SAFE(Scale for Adolescents) 척도를 수정하여 사용한 노충래(2000)의 문항을 재수정하여 사용한 홍진주(2004)의 척도를 사용하였다. 이 척도는 사회적, 태도적, 가족적, 환경적 문화적응스트레스를 측정하며 4점 Likert 척도로 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’ 4점 척도록 점수가 높을수록 문화적응 스트레스가 높음을 의미한다. 홍진주(2004) 연구에서의 Cronbach's  $\alpha$ 는 .76이었고 본 연구에서 도구의 Cronbach's  $\alpha$ 는 .862이었다.

### 2) 종속변수 : 학교생활

학교생활은 청소년종합실태조사 김정남(2013)의 척도를 사용하였다. 이 척도는 “수업 시간이 재미있다”, “공부하기에 좋은 분위기다”, “우리학교 선생님들이 좋다”, “우리학교 친구들이 좋다” 총 4문항으로 구성되었다. 각 문항은 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’ 4점의 Likert 척도로 점수가 높을수록 학교생활적응을 잘하는 것을 의미하며, 본 연구에서 도구의 Cronbach's  $\alpha$ 는 .756이었다.

### 3) 매개변수 : 친구의지지

친구의 지지는 한미현(1996)의 척도를 사용하였다. 이 척도는 “내 친구들은 나에게 관심이 많은 것 같다”, “내 친구들은 나와 함께 지내는 것을 좋아하는 것 같다”, “내 친구들은 나를 잘 이해해주는 것 같다” 총 3문항으로 구성되었다. 각 문항은 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’ 5점의 Likert 척도로 점수가 높을수록 친구의지지 정도가 높음을 의미하며, 본 연구에서 도구의 Cronbach's  $\alpha$ 는 .844이었다.

### 4) 매개변수 : 자아존중감

자아존중감은 Rosenberg (1965)의 도구에서 발췌하여 사용하였다. 이 척도는 ‘내가 다른 사람들처럼 가치있는 사람이라고 생각한다’, ‘좋은 성품을 가졌다고 생각한다’, ‘내 자신에 대하여 대체로 만족한다’ 총 3문항으로 구성되었다. 각 문항은 ‘전혀 그렇지 않다’ 1점에서 ‘매우 그렇다’ 4점의 Likert 척도로 점수가 높을수록 자아존중감이 높음을 의미하며, 본 연구에서 도구의 Cronbach's  $\alpha$ 는 .818이었다.

### 4. 자료 분석방법

연구대상자의 인구사회학적 특성을 확인하기 위하여 SPSS Statistics 26.0으로 기술통계분석과 상관관계를 분석하였으며, SPSS Macro Process 4.0 프로그램을 활용하여 이중매개효과 분석을 실시하였다.

### 3. 연구결과

Table 1. General Characteristics (N;2,271)

Variable	Categories	Frequency	%
성별	남	1265	51.5
	여	1006	48.5
거주지역	서울	232	10.2
	경인	590	26.0
	충청 및 강원	556	24.5
	경상권	533	23.5
	전라 및 제주	360	15.8
경제적수준	어주 어렵다	105	4.6
	어려운 편이다	529	23.3
	보통이다	1473	64.9
	잘사는 편이다	157	6.9
	아주 잘 산다	7	0.3

Table 2. Descriptive statistics and correlations of variables

Categories	M	SD	Range	학교생활적응	문화적응스트레스	친구의지지	자아존중감
학교생활적응	2.54	.39	1~4	1			
문화적응스트레스	3.36	.32	1~4	.211**	1		
친구의지지	2.92	.53	1~5	.391**	.170**	1	
자아존중감	2.50	.42	1~4	.327**	.178**	.396**	1

\*p<.05, \*\*p<.01, \*\*\*p<.001



Table 3. 대상자의 문화적응스트레스와 학교생활의 관계에서 친구의지지와 자아존중감의 매개효과

	Variables	b	SE	t	p	95% CI	
						LLCI	ULCI
	(Constant)	5.89	0.37	15.88	<.001	5.16	6.62
	문화적응스트레스→친구의지지	0.17	0.01	7.85	<.001	0.07	0.11
	R=.17, R2=.03, F=61.59, P=.001						
	(Constant)	3.37	0.29	11.78	<.001	2.81	3.93
	문화적응스트레스→자아존중감	0.11	0.01	5.61	<.001	0.03	0.09
Indirect effect	친구의지지→자아존중감	0.38	0.02	18.51	<.001	0.26	0.33
	R=.41, R2=.17, F=210.82, P=.001						
	(Constant)	3.83	0.36	10.79	<.001	3.14	4.53
	문화적응스트레스→학교생활적응	0.13	0.01	6.36	<.001	0.05	0.09
	친구의지지→학교생활적응	0.29	0.02	13.74	<.001	0.24	0.33
	자아존중감→학교생활적응	0.19	0.03	8.69	<.001	0.18	0.28
	R=.45, R2=.20, F=176.33, P=.001						
	(Constant)	6.69	0.36	18.82	<.001	5.60	7.39
Direct effect	문화적응스트레스→학교생활적응	0.21	0.01	9.83	<.001	0.09	0.14
	R=.21, R2=.04, F=96.76, P=.001						

b:regression coefficient

Table 4. Mediating effect verification

	변수	Effect	BootSE	95% CI	
				LLCI	ULCI
Total effect	Direct effect + Indirect effect	.08	.01	.06	.10
Direct effect	문화적응스트레스→학교생활	.05	.01	.03	.06
	문화적응스트레스→친구의지지→학교생활	.02	.01	.02	.04
	문화적응스트레스→자아존중감→학교생활	.01	.00	.01	.02
	문화적응스트레스→친구의지지→자아존중감→학교생활	.01	.00	0.01	0.02

## 골다공증 여성노인의 건강관련 삶의 질 영향요인 : 국민건강영양조사자료활용(2016~2020)

전은미<sup>1</sup>, 강창완<sup>2</sup>

### 요 약

본 연구는 골다공증 여성노인의 삶의 질에 영향을 미치는 요인을 확인하고자 2016~2020년 국민건강영양조사 자료의 골다공증 여성노인 1,679명을 대상으로 조사하였다. 수집된 자료는 SPSS WIN 25.0 프로그램으로 복합표본분석 방법을 이용하였다. 여성노인에서 인구사회학적 특성과 건강 관련 특성에 따른 골다공증 유병률의 차이는 Rao-Scott  $\chi^2$ 과 t-test로 분석하였고, 골다공증 여성노인의 일반적 특성에 따른 삶의 질의 차이는 t-test 또는 Wald F로 분석하였다. 골다공증 여성노인의 건강 관련 삶의 질에 영향을 미치는 요인은 일반선형회귀분석을 이용하였다. 연구결과 인구학적 특성과 건강관련 특성에 따른 골다공증 유병률은 연령( $\chi^2=5.53$ ,  $p=0.001$ ) 거주지( $\chi^2=11.42$ ,  $p=0.001$ ), 결혼상태( $\chi^2=7.31$ ,  $p=0.007$ ), 동거가족형태( $\chi^2=9.87$ ,  $p=0.001$ ), 경제상태( $\chi^2=4.84$ ,  $p=0.028$ ), 교육수준( $\chi^2=17.26$ ,  $p=0.001$ ), 우울증( $\chi^2=24.5$ ,  $p=0.001$ ), 지각된 건강상태( $\chi^2=44.67$ ,  $p=0.001$ ), 만성질환동반( $\chi^2=19.13$ ,  $p=0.001$ ), 행동제한( $\chi^2=555.6$ ,  $p=0.001$ ), 지각된 스트레스( $\chi^2=4.52$ ,  $p=0.034$ ), 음주( $\chi^2=5.98$ ,  $p=0.032$ ), 체질량지수( $\chi^2=7.49$ ,  $p=0.001$ ), 대사증후군( $\chi^2=9.82$ ,  $p=0.001$ )에서 통계적으로 유의한 차이가 있었다. 골다공증 여성노인의 건강관련 삶의 질은 연령( $F=18.61$ ,  $p=0.001$ ), 거주지( $t=13.40$ ,  $p=0.001$ ), 결혼상태( $t=21.31$ ,  $p=0.001$ ), 동거가족형태( $F=9.25$ ,  $p=0.001$ ), 경제상태( $t=18.13$ ,  $p=0.001$ ), 교육정도( $t=46.54$ ,  $p=0.001$ ), 경제활동( $t=4.13$ ,  $p=0.042$ ), 우울증( $t=17.39$ ,  $p=0.001$ ), 지각된 건강상태( $F=151.86$ ,  $p=0.001$ ), 만성질환동반( $t=31.31$ ,  $p=0.001$ ) 활동제한( $t=158.08$ ,  $p=0.001$ ), 지각된 스트레스( $t=61.17$ ,  $p=0.001$ ), 음주( $t=6.49$ ,  $p=0.011$ ), 체질량지수( $F=3.45$ ,  $p=0.032$ ), 대사증후군( $F=6.37$ ,  $p=0.002$ )에서 유의한 차이가 있는 것으로 나타났다. 골다공증 노인 여성의 건강관련 삶의 질에 영향을 미치는 변수는 동거가족 형태에서 독거인 경우, 가구소득이 평균 이상, 지각된 건강상태가 좋을수록, 우울증이 없는 경우, 행동제약이 없을 때, 지각된 스트레스가 없을수록, 나이와 체질량지수가 적을수록 삶의 질이 높았고 설명력은 34.1%였다.

주요용어 : 골다공증, 노인여성, 삶의 질, 주관적 건강상태, 우울.

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

<sup>2</sup>Professor, Industrial Management Big Data Engineering Major, Dongeui University, 176 Eomgwangno, Busanjin-gu, Busan, 47340, E-mail: cwkwang@deu.ac.kr

## 병기에 따른 관측중단 손실의 영향 감소 조정 DSM을 이용한 NSCLC 생존분석\*

박소연<sup>1</sup>, 강세령<sup>2</sup>, 김민수<sup>3</sup>

### 요약

폐암은 국내 사망률이 가장 높은 암으로 초기 증상이 폐에 국소 될 경우 5년 생존율이 76.7%이지만, 원격 전이 시 생존율이 11.5%로 감소한다. 폐암의 17.0%만이 조기 진단되므로 명확한 생존 진단 및 예후 분석은 이후 치료 방침 결정에 매우 중요하다. 본 연구에서는 화순전남대학교병원에서 제공된 NSCLC(Non-small cell lung cancer) 환자 2,687명의 임상 데이터를 이용한 생존분석을 통해 생존율을 예측한다. 임상 데이터는 수집 시점 및 의료기관 변경 등의 원인으로 오른쪽 관측 중단된 생존시간 데이터가 다수 존재한다. 관측중단 데이터가 많을수록 사용할 수 있는 정보를 감소시키고 편향이 증가하여 생존함수의 추정 성능을 저하한다는 문제가 발생한다. 이를 해결하기 위해 DSM(Deep Survival Machines) 모형의 관측중단 및 이벤트 발생 데이터의 결합된 loss를 조정하였다. 개별 변수 영향력이 높은 Overall stage를 기준으로 상한을 설정하여, 조건에 따라 관측중단 손실의 영향을 직접적으로 감소하도록 설정하였다. 생존율 예측 결과 다른 모형(CoxTime, DeepSurv, DCM(Deep Cox Mixture))에 비하여 C-index, Brier Score를 동일하게 유지하면서, MRL(Median Residual Life) 방법으로 결정된 MAE 값을 평균 44일 이상 적게 나타내어 우수한 성능을 보였다. 본 연구 모형은 3,200일 이후 Brier score가 꾸준히 감소하고 생존율 곡선에서 실제 관측중단 시점 이후 생존율이 꾸준히 감소하는 양상을 보인다는 특성을 가진다. 향후 본 연구를 통해 환자의 병기를 활용한 정확한 실제 생존율 예측이 환자 개개인의 예후 진단에 유용하게 활용될 수 있을 것으로 기대된다.

주요용어 : 생존분석, 생존율 예측, DSM, 관측중단, 편향 감소.

\* 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단 바이오·의료기술개발사업의 지원을 받아 수행된 연구임(NRF-2019M3E5D1A02067961).

<sup>1</sup>61186 광주광역시 북구 용봉로 77, 전남대학교 수학과/통계학과 석사과정. E-mail: syeonon1512@gmail.com

<sup>2</sup>58128 전라남도 화순군 화순읍 서양로 322, 화순전남대학교병원 핵의학과 교수.

E-mail: campanella9@naver.com

<sup>3</sup>(교신저자) 61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 교수. E-mail: kimms@chonnam.ac.kr

## 디퍼러닝 연계 교양교육튜터링 효과성 분석\*

김태영<sup>1</sup>, 윤혜경<sup>2</sup>, 최승배<sup>3</sup>

### 요약

디퍼러닝(deeper learning)은 학생들이 길러야 할 역량군 또는 달성해야 할 목표군을 기술하기 위해 사용되어 왔고, 이와 같은 목적에 도달하기 위한 학습법(또는 학습과정)의 특징을 나타내는 용도로 언급된다. 디퍼러닝의 목표는 “교과 내용을 명쾌하게 이해하고 이를 교실 내외 직업 현장에 적용할 수 있는 능력”을 향상시키는 것이며, 관련 연구에서는 디퍼러닝을 향후 마주하게 될 대학(college)생활, 직장(career)생활, 사회(civic)생활에서의 성공을 위해 필수불가결한 것으로 보고 있다. D 대학교 교양교육연구소는 2020년 1학기부터 지속적으로 교양교육튜터링 프로그램을 운영해오고 있다. 디퍼러닝이 추구하는 가치와 많은 부분을 공유하고 있는데, 디퍼러닝의 주요 요인인 자기조절학습, 의사소통능력 및 문제해결능력은 그간 운영해 온 교양교육튜터링 프로그램의 취지와도 부합한다. 이러한 관점에서, 본 연구에서는 교양교육튜터링을 통한 디퍼러닝의 수준 변화를 측정하고 이를 통해 프로그램의 효과성(effectiveness)을 계량화하고자 한다. 본 연구는 교양교육튜터링 프로그램 참여자들과 비참여자들의 디퍼러닝 수준을 비교하였다. 비교 결과가 추후 제시될 예정이며, 이와 관련하여 고등교육 현장에서 특정 프로그램의 교육효과성 논의가 추가될 것이다.

주요용어 : 디퍼러닝, 표면적 학습, 교양교육튜터링, 교육효과성, 공분산분석.

\*이 논문은 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2022S1A5C2A04093562).

<sup>1</sup>(교신저자) 47340 부산시 부산진구 엄광로 176, 동의대학교 디그니타스교양교육연구소, 연구교수.  
E-mail: tkim33@deu.ac.kr

<sup>2</sup>47340 부산시 부산진구 엄광로 176, 동의대학교 기초과학교양학부, 교수. E-mail: hgyoon@deu.ac.kr

<sup>3</sup>47340 부산시 부산진구 엄광로 176, 동의대학교 산업경영·빅데이터공학전공, 교수. E-mail: csb4851@deu.ac.kr

## cs-QCA를 활용한 환경분쟁조정 인과성 분석: 층간소음 사례연구\*

송경렬<sup>1</sup>

### 요약

이 연구는 QCA 연구 방법을 탐색한다. QCA는 집합이론에 따른 필요·충분조건의 논리에 따라 비대칭관계를 이용한다. 독립변인인 원인 조건들이 서로 결합하여 항(term)이나 식(solution)을 구성하면, 원인 조건의 결합을 통하여 결과 조건에 대한 설명력을 더욱 높일 수 있는 장점이 있다. 이러한 점은 원인 조건과 결과 조건의 대칭적 선형관계를 파악하는 회귀분석과 비교되기도 한다. 예시로서 이 연구에서는 QCA 연구 방법에 따라 일상적 환경분쟁으로 꼽히는 층간소음 47개 사례를 분석한다. 분쟁조정 메커니즘에 따라 도출된 원인 조건과 결과 조건을 비대칭관계로 표현하면 다음과 같다. '(당사자 유형, 분쟁 경과 기간, 사실조사, 인과관계) → Y(당사자 합의여부)'. 분석 결과 도출된 연구모형에서는 원인 조건의 결합양상에 따라 4가지 경로가 나타난다. 원인 집합의 부분집합 정도를 나타내는 일치성은 0.921이며, 원인 집합이 결과 집합을 설명하는 정도인 포괄성은 0.972이다. 이는 원인 조건 4가지가 결합하면 97.2%의 설명력을 갖는다는 것을 의미한다. 분석에 따른 해석적 의의는 다음과 같다. 인과성 분석에 있어서 QCA 연구 방법은 집합이론에 따른 비대칭관계를 활용한다. 이러한 점은 결과변수의 순효과에 초점을 맞추는 변수지향적 회귀분석에서 나타나는 대칭관계의 설명력을 상호보완할 수 있는 장점이 된다.

주요용어: cs-QCA, 원인 조건 결합(conjunctural causation), 비대칭관계, 인과성(causality).

### 1. 서론

인과성 분석의 변수 혹은 변인 간 관계에서 '=' 과 '< or >'의 차이점은 무엇일까?

이 연구는 방법론적 관점에서 집합이론을 이용한 cs-QCA(crisp-set qualitative comparative analysis: 이하 cs-QCA)에 따른 인과성 분석을 시도한다. 구체적으로 일상생활의 대표적 환경분쟁인 층간소음 사례연구를 통하여 이 분쟁 해결의 메커니즘에는 어떤 원인 조건과 결과 조건이 있으며, 이러한 원인 조건은 어떤 결합양상으로 나타나는가를 살펴본다. 질적 비교분석인 QCA는 원인 조건과 결과 조건의 관계뿐만 아니라, 원인 조건들이 결합하는 양상에 대한 설명력을 가진다.

QCA의 가장 큰 특징은 집합이론에 따른 비대칭관계에 관한 연구 방법이다. 독립변인인 원인 조건들이 서로 결합하여 항(term)이나 식(solution)을 구성한다. 이 점에 있어서 원인 조건과 결과 조건의 상관적 대칭관계를 파악하는 회귀분석과 비교되기도 한다. 따라서 이 연구는 인과성 분석의 관점에서 대칭관계와 비대칭관계 간의 차별성과 상호보완성을 설명함에 의의가 있다.

\*이 연구는 2019년 서울대 행정논총 게재논문을 KDAS 발표목적에 맞게 편집함(ISSN 1229-6694).

<sup>1</sup>02450 서울 동대문구 이문로 107 한국외국어대학교 국정관리연구소 연구원. E-mail: marsmas@hufs.ac.kr

## 2. QCA를 활용한 연구 방법

QCA는 집합이론을 활용하여 인과관계를 설명한다(Ragin, 2000; 이승윤, 2014)<sup>34</sup>. 사회과학 명제는 대부분 집합론적인 속성을 갖는 경우가 많다. 따라서 원인과 결과 간에 집합관계가 존재한다면, 사회현상에서 나타나는 인과성을 필요조건과 충분조건이라는 집합이론으로 설명할 수 있다.

### 2.1. QCA: 비대칭관계로서의 인과성 분석

원인 조건과 결과 조건이 집합관계로 파악되면 비대칭관계에서 인과성을 분석(안건훈, 2005; Schneider and Wagemann, 2012)해야 한다. 반면 회귀분석 전제조건은 독립변수가 변화하면 종속변수의 방향과 영향의 강도도 변화한다는 점이다. 원인이 존재하면 결과도 존재하고, 원인이 부재하면 결과도 부재해야 한다. 양적방법론의 상관관계에서 인과성을 설명하는 경우, 원인 조건과 결과 조건은 완벽히 대응되어야 한다(이승윤, 2014). 그런데 독립변수와 종속변수 간의 관계가 유의미하다고 하여 반드시 인과관계가 성립하지는 않는다. 원인이 부재하는데 결과가 존재하는 경우 상관관계로 설명할 수 없다. 이러한 현상은 집합관계에서 잘 나타난다. 따라서 집합관계가 존재하는 경우 필요조건 혹은 충분조건의 논리로 인과관계를 파악해야 한다. QCA는 사례의 속성을 결합하여 총체적 맥락(configurative context)으로 해석한다. 동일한 결과가 발생하는 동귀결성(equifinality), 즉 다양한 원인 조건의 결합(conjunctural causation)이 있음을 강조한다. 또한 전체집합과 부분집합이라는 집합론적 특성에 따라 원인과 결과 간의 비대칭성(causal asymmetry)이라는 특징이 있다(Ragin, 2000). 이러한 점은 선형 분석에서 인과관계의 결정변수가 무엇인가에 따라 결과변수의 순 효과(net effect)에 초점을 맞추는 변수지향적 접근과 크게 비교된다.

### 2.2. cs-QCA: 원인의 존재와 부재

cs-QCA는 집합대수(Boolean Algebra)로 인과관계 조건의 유무, 즉 존재(1) 혹은 부재(0)라는 이항 형태를 띠게 된다. 이러한 이분적 사고는 일상생활에서 자주 나타난다(Schneider & Wagemann, 2012: 59). cs-QCA 첫 단계는 원인 조건과 결과 조건으로 분류된 사례를 '1'과 '0'으로 코딩한 후, 진리표(Truth Table)로 구성하는 것이다. 진리표에는 논리적으로 가능한 원인변수의 조합만큼의 행(줄)을 갖게 되며, 이때 조합의 수<sup>35</sup>는  $2^k$ 가 된다. 중요한 점은 이러한 조합들이 경로(path)의 형태로서 그룹화( $3k-1$ )<sup>36</sup>가 된다는 점이다. cs-QCA 조합논리는 사례를 변수로 분해하지 않고, '전체적 맥락'에서 본다. 특정 원인만을 독립변인으로 간주하지 않고, 다른 원인 조건의 존재와 부재의 맥락에서 파악한다. 만약 결과 E에 대하여 원인조건 A,  $\sim B$ ,  $\sim C$ 라는 3가지 원인조건들이 결합하여  $A \cdot \sim B \cdot \sim C \rightarrow E$ 라는 해결식(solution)이 도출되면, 조합논리 관점에서는 결과 E가 A에 의해서만 발생하였다고 해석하지 않는다. 결과 E는 A라는 원인 조건과 더불어 B와 C가 부재한 경우 발생한다고 해석한다.

<sup>34</sup> QCA의 종류는 cs-QCA, fuzzy-set QCA, multi-value QCA, temporal QCA 등이 있다.

<sup>35</sup> 원인조건이 3개인 2진 독립변수라면, 조합 수는 '0, 1, 0', '1, 1, 0' 같이 8개 행을 갖는 조합이 구성된다.

<sup>36</sup> a, b, c라는 3가지 속성을 가진 조합 수는 8행이다. 그룹화가 이루어지면, 1가지 속성 조합(a,  $\sim a$ , b), 2가지 속성 조합(a·b,  $\sim a \cdot b$ , b·c), 3가지 속성 조합(a·b·c,  $\sim a \cdot b \cdot \sim c$ ) 등 26개 그룹이 된다. 조합들은 경로 형태를 「 $\sim a + \sim a \cdot b + b \cdot c + \sim a \cdot b \cdot \sim c \rightarrow e$ 」와 같은 해결식(solution)을 구성한다.

### 3. 분석 및 해석: 층간소음 분쟁사례

연구 대상인 관찰단위는 1992~2016년까지 중앙·지방환경분쟁조정위원회에서 공개한 47개의 층간소음 분쟁사례이다. 시간적 경로를 통한 사례 내 분석과 집합이론의 최소화 논리를 통하여 도출한 원인 조건과 결과 조건 간의 모형 특성은 Figure 1과 같다.

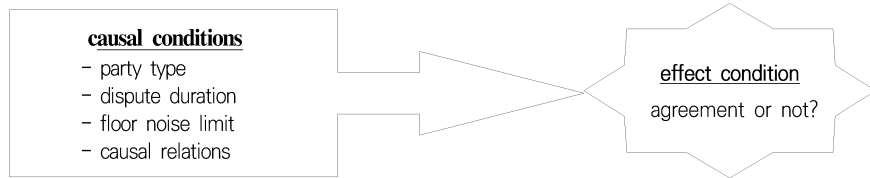


Figure 1. Model Specification

#### 3.1. 진리표 분석(truth table analysis)

cs-QCA는 집합대수를 사용하여 진리표를 분석한다. 인과관계를 집합이론으로 파악하기 때문에 필요조건과 충분조건에 대한 설명이 필요하다. 이때 인과관계 설명력을 높이기 위한 보정계수(parameter fit)가 활용된다. 일치성(consistency)<sup>37</sup>과 포괄성(coverage)<sup>38</sup>이 그것이다. 이 두 기준은 원인 집합이 결과 집합을 어느 정도 설명하는가를 파악하는데 주요 역할을 한다. 원인 조건이 4개이면 조건 결합은 24=16개 행(row)으로 나타난다. 47개 층간소음 분쟁사례에서는 9개 행이며, 0.8 이상의 일치성을 보이는 결합조건 열(column)은 1~7번까지다. 7개 행은 원인 조건결합이 결과 조건에 대하여 설명력이 있음을 의미한다. 일치성이 0.8 이상일 경우 원인 집합은 결과 집합에 거의 언제나(almost always) 충분조건이 된다(Ragin, 2006). 일정부분이 생략(2~6)된 진리표는 Table 1과 같다.

Table 1. Truth Table

Conjunctural type	Causal conditions				Effect Condition	Cases	Consistency
	Party Type	Dispute Duration	Noise Limit	Causal Relation			
1	0	0	0	0	1	4	1
7	1	0	0	0	1	8	0.875
8	0	1	1	1	0	7	0.142857
9	0	0	1	1	0	2	0

#### 3.2. 해결식(solution formula)에 따른 해석

진리표 분석결과가 해결식(solution formula)이다. 이는 원인 조건결합에 따라 최소화 논리(minimization logic)에 의거, 3가지 모형<sup>39</sup>으로 나타난다. 3가지 모형은 결합을 구성하는 경로(path)

<sup>37</sup> 원인 집합이 결과 집합의 부분집합인 정도(%), 0.75 이상이면 인과성 설명력이 높다는 것을 의미한다.

<sup>38</sup> 원인 집합이 결과 집합을 설명하는 정도(%), 원인 집합과 결과 집합 간의 크기(size)를 비교한다.

<sup>39</sup> 원인 조건결합의 최소화 정도에 따라 결합구성 해결식은 복잡 모형(conservative solution), 간결 모형(parsimonious solution), 중간 모형(intermediate solution)으로 나뉜다. 집합 이론적 관점에서 세 모형은 ‘복잡 모

양상이 다르게 나타날 수 있다. 이 연구에서는 중간 모형(intermediate solution)의 분석 결과를 해석한다. 중간모형은 표준분석에 가장 근접하는 모형(Ragin, 2017: 42)이며, 제한된 다양성(limited diversity)<sup>40</sup>이라는 현상의 설명 공간을 제공한다. Table 1의 분석 결과에 따른 중간모형이 Table 2이다.

Table 2. Intermediate Solution<sup>41</sup>

Conjunctural condition	Raw coverage	Uni. coverage	Consistency
Frequency cutoff: 1, Consistency cutoff: 0.875			
① ~NI * ~CR	0.833333	0.555556	0.909091
② PT * ~DD * ~ CR	0.222222	0.0277778	0.888889
③ ~PT * DD * ~CR	0.166667	0.0833333	1
④ PT * DD * NI * CR	0.0277778	0.0277778	1
Solution coverage: 0.972222, Solution coverage: 0.921053			

Table 2 중간모형을 비대칭관계로 표현하면, f(당사자 유형, 분쟁경과기간, 사실조사, 인과관계) → Y(당사자 간 합의 여부)이다. 중간모형에서는 결합조건에 따른 4가지 경로(paths)를 갖는다. 여기서 원인 집합이 부분집합의 정도를 나타내는 전체 일치성(solution consistency)은 0.921이며, 원인집합이 결과집합을 설명하는 정도인 전체 포괄성(solution coverage)은 0.972이다. 이는 원인 조건 4가지가 결합하면 97.2%의 설명력을 갖는다는 것을 의미한다. 이를 일반 언어형식으로 변환하면 다음과 같다.

「f(① + ② + ③ + ④) → Y」

①이거나(or), ②이거나(or), ③이거나(or) ④이면, 분쟁해결 가능성이 높다.

즉 ‘①소음 수인한도 이내이고, 소음 개연성이 없거나, ②당사자 유형이 개인 대 개인이고, 분쟁기간이 짧고, 소음 개연성이 없거나, ③당사자 유형이 개인 대 개인이 아니고, 분쟁기간이 짧지 않고, 소음개연성이 없거나, ④당사자 유형이 개인 대 개인이고, 분쟁기간이 짧지 않고, 소음 수인한도를 초과하고, 소음 개연성이 있다면, 당사자간 합의에 따른 분쟁해결 가능성이 높다.’고 해석할 수 있다.

## References

- Ahn, G. H. (2005). *An Analysis of Causality*. Seoul: SNU Press. (in Korean).
- Lee, S. Y. (2014). *Theory & Application of Fs/QCA*: KU Press. (in Korean).
- Marx, Axel. (2010). Crisp-set qualitative comparative analysis and model specification: Benchmarks for future csQCA applications. *International Journal of Multiple Research Approaches*, 4(2): 138-158.
- Ragin, Charles C. (2017). *User's Guide to Fuzzy-Set/Qualitative Comparative Analysis*. University of California Press. cragin@uci.edu.----- (2000). *Fuzzy Set Social Science*. Chicago: University of Chicago Press.
- Schneider, Carsten. Q., & Wagemann, Claudius. (2012). *Set-Theoretic methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge University Press.
- Song, K. Y. (2019). *An Analysis of Causality: Resolutions of Noise between Floors Disputes*. <http://dx.doi.org/10.24145/KJPA.57.4.8>. (in Korean).

형 ⊂ 중간 모형 ⊂ 간결 모형'의 집합관계가 형성되어 간결 모형이 상위집합으로서 가장 큰 포괄성을 갖는다.

<sup>40</sup> 이는 논리와 실제 간의 간극이 생긴 것으로 논리적 잔차로 간주된다(Schneider and Wagemann, 2012: 328).

<sup>41</sup> Table 2의 항목은 PT(party type), DD(dispute duration), NL(noise limit), CR(causal relation)으로 약칭한다.



## 교통 네트워크 데이터에 대한 거대언어모델 활용 방안 연구

정재경<sup>1</sup>, 김예은<sup>2</sup>, 이다운<sup>3</sup>, 최호식<sup>4</sup>

### 요약

최근 거대언어모델이 자연어처리 이외의 영역에서도 우수한 성능을 보이며 범용성이 확대됨에 따라 이를 도시 현안 관리에 적용하려는 시도들이 이루어지고 있다. 본 연구는 거대언어모델을 교통 네트워크 시스템 관리에 활용할 수 있는 방안을 탐색해 보는 것을 목표로 한다. 표준화된 전국 교통 네트워크 데이터와 개별차량의 출발지/도착지(O/D), 이동 경로 시퀀스를 거대언어모델이 이해할 수 있는 문장 형식으로 구성하여 교통 네트워크 데이터를 이해하는 거대언어모델을 구축한다. 개별차량의 GPS 경로 데이터 자료구조에 적합한 토큰화 단위를 설계하고 문장에 조합하여 GPT 기반의 polyglot-KO 1.3B 한국어 모델을 학습한다. 또한, 교통 도메인 관련 다양한 다운스트림 태스크를 수행하기 위해 추가적인 프롬프트 엔지니어링 방법론을 고찰한다. 교통 네트워크 데이터를 학습한 거대언어모델을 NLP 기반 질의 태스크에 적용하여 이의 실증결과와 향후 유효성을 살펴본다.

주요용어 : 거대언어모델, 교통 네트워크, Prompt-learning.

---

<sup>1</sup>02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 박사과정.

E-mail: jaekyeong@uos.ac.kr

<sup>2</sup>02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 석사과정.

E-mail: keeun04@uos.ac.kr

<sup>3</sup>02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 석사과정.

E-mail: dlekdns07@uos.ac.kr

<sup>4</sup>(교신저자) 02504 서울특별시 동대문구 서울시립대로 163, 서울시립대학교 도시빅데이터융합학과 교수.

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

## Robust Graph Augmentation for Molecular Property Prediction

*Soyoung Cho*<sup>1</sup>, *Sungchul Hong*<sup>2</sup>, *Changwon Lim*<sup>3</sup>, *Jong-June Jeon*<sup>4</sup>

### Abstract

Graph Neural Networks (GNNs) exhibit potential in predicting the properties of molecules, but computational analyses with the GNNs often encounter the problem of data imbalances. Augmentation techniques have emerged as a popular solution to this problem, and they enhance the model's generalization capacity. Common augmentation approaches include graph regularizers that alter graphs' structure or adversarial perturbation to node features. This study proposes an adversarial perturbation on influential graph features to reduce its computation cost. Our method is based on approximating the influence function that measures the importance of each training sample on prediction. We approximate the influence function by the limited memory BFGS (LBFGS) algorithm during the model training process. In order to assess the efficacy of different approaches in approximating influence function, we conduct experiments that demonstrated the effectiveness of leveraging influential samples compared to using the entire dataset.

*Keywords* : Graph Neural Network, Influence Function, Data Augmentation, Molecular Property Prediction.

---

<sup>1</sup>02054 Graduate Student, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Korea.  
E-mail: sy.cho@uos.ac.kr

<sup>2</sup>02054 Postdoctoral Researcher, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Korea.  
E-mail: shong@uos.ac.kr

<sup>3</sup>06974 Professor, Department of Applied Statistics, Chung-Ang University, Heukseok-ro 84, Korea.  
E-mail: clim@cau.ac.kr

<sup>4</sup>(Corresponding Author) 02504 Professor, Department of Statistics, University of Seoul, Seoulsiripdaero 163, Korea. E-mail: jj.jeon@uos.ac.kr

## Sparse Sufficient Dimension Reduction via Outer Product of Gradients using Composite Quantile Regression

*Seogyoung Lee*<sup>1</sup>, *Seung Jun Shin*<sup>2</sup>

### Abstract

Sufficient dimension reduction (SDR) has been a steady choice for reducing the dimensionality of predictors by finding the central subspace, defined as an intersection of subspace of predictors that contains all the regression information. When the number of predictors is large, it is natural to assume that only a small subset of predictors is informative. In this context, sparse SDR is desirable approach, which perform variable selection and dimension reduction at the same time. In this article, we propose the sparse outer-product of gradients (OPG) estimator to recover the central subspace via directly estimating the gradients of multiple quantile regression functions. We demonstrate the promising performance of proposed method by simulations and real data analysis.

*Keywords* : Sparse Sufficient Dimension Reduction (sparse SDR), Gradient Learning, Outer-Product-of-Gradients (OPG), Composite Quantile Regression, Kernel Method.

---

<sup>1</sup>Graduate Student, Department of Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: syl0101@korea.ac.kr

<sup>2</sup>(Corresponding Author) Associate Professor, Department of the Statistics, Korea University, Anam-ro, Seongbuk-gu, Seoul, 02841, Korea. E-mail: sjshin@korea.ac.kr

## 양상블 모델을 이용한 MMORPG의 봇 탐지 연구

정현수<sup>1</sup>, 최정원<sup>2</sup>, 원상준<sup>3</sup>, 김수인<sup>4</sup>, 정윤서<sup>5</sup>

### 요 약

온라인 대규모 다중 사용자 온라인 역할수행 게임(MMORPG)의 인기가 상승함에 따라, 자동화된 게임 봇을 운영하는 작업장(Gold Farmer Group)이 게임 생태계에 더욱 강력한 영향력을 행사하고 있다. 게임 봇들은 일반 게임유저들의 피로도를 증가시키며 인게임 시장에 혼란을 초래하기에, 올바른 봇 탐지가 중요하다. 선행 연구와 게임 회사의 기존 봇 탐지 모델들은 주로 게임 내 행동분석과 규칙 기반 시스템으로, 작업장의 게임 봇이 일반 유저의 게임 행동을 모방하는 방향으로 끊임없이 변화하는 상황에 유연하게 대응하기 어렵다. 본 연구에서는 일반 게임유저와 행동 패턴이 유사하더라도, 지속적, 효율적인 게임 자산 축적과 현금 거래 수행을 목표로 한다는 점에 착안하여 새로운 봇 탐지 모델을 제시한다. 아이템 시세 대비 비정상적인 아이템 거래 관련 변수들과 재화 보유량 상태의 시계열적 패턴을 사용한 MMORPG 봇 탐지 모델을 제시하여 기존 봇 탐지 모델들의 한계점을 개선한다. 사전에 게임 봇으로 탐지된 유저정보를 분류 범주(label)로, 재화 보유량 상태의 시계열적 패턴을 학습하는 LSTM(Long Short Term Memory)과 아이템 거래 관련 변수들을 학습하는 랜덤 포레스트 모형을 결합한 양상블 모델을 제안한다. 카카오 게임즈 오딘 게임 로그 데이터를 기반으로 잠재적인 게임 봇들을 탐지하고, 모델이 봇으로 예측한 유저들과 일반 게임유저들의 차이를 확인하였다. 마지막으로, 이미 작업장으로 탐지된 유저들과 유사한 특성을 가지지만 아직 탐지되지 않은 유저들을 게임 봇 유저 후보군으로 제시하였다.

주요용어: 게임 봇 탐지, 딥러닝, 양상블 모델, 재화 상태 패턴.

<sup>1</sup>02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 석사과정. E-mail: hs5042@korea.ac.kr

<sup>2</sup>02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 석사과정. E-mail: jung1232@korea.ac.kr

<sup>3</sup>02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 석사과정. E-mail: kurum279@korea.ac.kr

<sup>4</sup>02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 석박통합과정. E-mail: ksi2002@korea.ac.kr

<sup>5</sup>(교신저자) 02841 서울특별시 성북구 안암로 145, 고려대학교 통계학과 교수 E-mail: yoons77@korea.ac.kr

## Functional adaptive group lasso with its non-asymptotic bounds<sup>\*</sup>

*Sehun Jang*<sup>1</sup>, *Jun Song*<sup>2</sup>

### Abstract

In this manuscript, we develop theories and methods for functional predictor selection in multivariate functional data in a scalar-on-function regression problem. Current existing methods either lack valid theoretical validation or require overly strong assumptions that cannot be verified using the sample. To address this issue, we construct a novel distance measure for infinite dimensional data and introduce a new penalty scheme for functional data to derive superior asymptotic properties under more relaxed and reasonable assumptions. Moreover, we examine the non-asymptotic behavior of this new method under a finite-sample design and derive its non-asymptotic bounds to demonstrate selection and estimation consistency. Simulation results indicate the method's effective performance, and it is anticipated that the novel distance-based penalty scheme can enhance numerous other penalized methods for functional data.

*Keywords* : functional data analysis, adaptive group lasso, function selection, consistency criteria, non-asymptotic bound.

---

<sup>\*</sup>For Jun Song, this work was supported by the National Research Foundation of Korea(KRF) grant funded by Korea government(MSIT) (No. 2022RIC1003647).

<sup>1</sup>Department of Statistics, Korea University. E-mail: sahoon1004@korea.ac.kr

<sup>2</sup>(Corresponding Author) Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, 02841, South Korea. E-mail: junsong@korea.ac.kr

## 일반화된 지수 분포를 고려한 NHPP 소프트웨어 신뢰성 모형에 관한 연구\*

송광윤<sup>1</sup>, 장인홍<sup>2</sup>, 김윤수<sup>3</sup>, 이다혜<sup>4</sup>

### 요 약

4차 산업혁명 시대에는 대부분의 기술을 구현하는 데 소프트웨어가 사용됩니다. 4차 산업혁명 시대에 소프트웨어는 매우 중요한 요소가 되고 있으며, 다양한 곳에서 소프트웨어가 활용되고 있습니다. 일반적으로 소프트웨어 시스템은 사용 기간이 길어질수록 성능이 저하되고 고장률이 증가하는 것으로 알려져 있습니다. Goel, Okumoto은 비동질성 포아송 과정(NHPP)에 기반한 지수 곡선을 고려한 소프트웨어 신뢰성 모형을 소프트웨어 신뢰성 모형 중 가장 먼저 제안하였고, Yamada et al.은 결함 검출률이 원래 시스템에 남아 있는 결함 수와 불완전한 디버깅으로 인해 새로 발생한 결함 수의 합에 비례한다고 가정하여 불완전한 디버깅을 적용한 소프트웨어 신뢰성 모형을 제안했습니다. 제어된 테스트 환경 내에서 결함 강도를 함수와 평균값 함수를 통해 소프트웨어의 잔여 결함 수, 고장률 및 신뢰성과 같은 신뢰성 지표를 추정하기 위해 많은 NHPP 소프트웨어 신뢰성 모델이 개발되었습니다. 일반적으로 이러한 모델은 소프트웨어 테스트 데이터에 적용된 다음 소프트웨어 운영 환경에서 소프트웨어 오류 및 신뢰성을 예측하는 데 사용됩니다. 그러나 소프트웨어 시스템은 다양한 위치에서 사용될 수 있습니다. 최근에는 소프트웨어 운영 환경의 불확실성을 고려한 다양한 고장 검출률 함수를 적용한 NHPP 소프트웨어 신뢰성 모델이 제안되고 있습니다. 일반화된 지수(GE) 분포는 지수 분포와 유사하지만 항상 모양 매개변수에 따라 위험률이 증가하거나 감소합니다. 또한 GE 분포는 감마 분포의 특성과 유사하지만 분포 함수는 감마 분포에 비해 계산이 간단하다는 장점이 있습니다. 일부 연구자들은 지수화된 각 분포에 대해 몇 가지 수학적 특성을 제시하기도 했습니다. 본 논문에서는 NHPP에 기반한 감마 분포를 이용하여 운영 환경의 불확실성을 고려한 일반화된 지수(GE) 고장 검출률 함수를 갖는 새로운 소프트웨어 신뢰성 모형을 제안하고, 제안된 모형이 다른 모형에 얼마나 우수한지 적합도를 통해 비교하고자 한다. 비교 결과, 새롭게 제안된 모형이 다른 모형에 비해 모든 척도의 값이 가장 낮게 나타나 유의미하게 우수함을 확인할 수 있습니다.

주요용어 : 감마분포, 결함 탐지율 함수, 비동질성 포아송 과정, 소프트웨어 신뢰성 모형, 지수 분포.

\*이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2021R111A1A01059842, NRF-2021R1F1A1048592, NRF-2021R1A6A3A01086716).

<sup>1</sup>61452 광주광역시 동구 필문대로 309, 조선대학교 컴퓨터통계학과, 조교수. E-mail: csssigg@chosun.ac.kr

<sup>2</sup>(교신저자) 61452 광주광역시 동구 필문대로 309, 조선대학교 컴퓨터통계학과, 교수.

E-mail: ihchang@chosun.ac.kr

<sup>3</sup>61452 광주광역시 동구 필문대로 309, 조선대학교 대학원 전산통계학과, 박사수료.

E-mail: imk92315@naver.com

<sup>4</sup>61452 광주광역시 동구 필문대로 309, 조선대학교 컴퓨터통계학과, 박사후연구원. E-mail: ec4869@naver.com

## 확률 과정 기법을 이용한 사육 환경 변화에 따른 한우 증체량 예측\*

강소라<sup>1</sup>, 조완현<sup>2</sup>, 나명환<sup>3</sup>

### 요약

농장에서는 한우의 무게를 통해 건강, 성장 상태 등을 확인하고 한우의 상태에 따라 사료량을 조절한다. 하지만 무게를 측정하거나 사료량을 조절하는 것은 어려운 일이며, 사육 환경에 따라 한우의 성장이 저해될 수 있다. 사육 환경은 스트레스를 유발하는 주요 요인으로 고온기에는 사료 섭취량과 산유량이 감소하여 체중이 절반 이상 감소하고 심하면 폐사한다. 따라서 사육 환경은 한우 성장에 중요한 요소이며 적절한 사육 환경을 유지하는 것이 필요하지만 한우 성장과 환경정보와의 관계성을 분석하는 연구가 많지 않다. 본 연구에서는 기상환경을 이용하여 한우의 증체량을 예측하는 모델을 제시하고자 한다. 먼저, 상관분석을 통해 한우 무게와 환경요인과의 관계를 규명하고, 일반적인 통계적 모형인 가중 선형 회귀모형과 확률과정을 기반한 가우시안 프로세스 모형을 통해 사육 기간 동안의 한우 무게 변화를 예측하고자 한다. 또한, 패널 데이터의 특성을 고려하여 가우시안 프로세스 패널 회귀모형을 비교 분석하고자 한다. 궁극적으로 한우 성장에 영향을 주는 중요한 요인을 알아내고 정밀한 사양 관리 의사결정을 지원하는 시스템을 제공할 것으로 기대된다.

주요용어 : 한우, 증체량 예측, 가중회귀모형, 확률과정모형, 가우시안 프로세스 패널 모형.

### 1. 서론

한우는 인기 있는 고기 중 하나이며, 적절한 사육 기간 동안의 한우 무게에 관심이 많다. 농장에서는 한우의 상태에 따라 사료량을 조절하며 각 한우의 무게를 살피는데, 건강이나 성장 상태를 확인하고 적절한 출하 무게에 도달하였는지 확인하기 위해 중요한 작업이다. 하지만 가축의 사료 섭취량을 조절하는 것과 가축의 무게를 측정하는 것은 어려운 일이며, 그 과정에서 많은 스트레스를 유발하여 성장이 저해될 수 있다. 스트레스를 발생시키는 요인에는 사육 환경의 영향이 크며, 축사에서는 온도와 상대습도만으로 산정된 가축 더위지수를 측정하여 예보하고 있다. 더위지수가 상승하면 사료 섭취량과 산유량이 감소하여 체중이 절반 이상 감소하기도 하고 심할 경우 폐사하는 지경에 이른다. 특히 비육우의 경우 기온이 30℃ 이상이 되면 발육이 정지된다. 농장에서는 여러 가지 비타민들을 사료에 첨가하여 한우에게 먹임으로써 고온 스트레스를 줄이고 있다. 하지

\*이 연구는 농림축산식품부 및 과학기술정보통신부, 농촌진흥청의 재원으로 농림식품기술기획평가원과 재단법인 스마트팜연구개발사업단의 스마트팜다부처패키지혁신기술개발사업의 지원(421017-04)과 교육부 및 한국연구재단의 4단계 두뇌한국21 사업 (4단계 BK21 사업, NO.5120200913674)으로 지원된 연구임.

<sup>1</sup>(제1저자) 61186 광주광역시 북구 용봉동 용봉로 77, 전남대학교 수학과/통계학과 박사과정.

Email: sc12love@gmail.com

<sup>2</sup>61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 교수. Email: whcho@chonnam.ac.kr

<sup>3</sup>(교신저자) 61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 교수. Email: nmh@chonnam.ac.kr

만 고온으로 인해 사료의 소화율이 떨어지게 되고, 맥박수, 호흡수, 체온도 크게 변화하게 된다. 따라서 사육 환경은 한우 성장에 중요한 요소이며 적절한 사육 환경을 유지하는 것이 시급하지만 한우 성장과 기상환경과의 관계성을 분석하는 연구가 많지 않다.

본 연구에서는 상관분석을 통해 한우의 무게와 기상 환경과의 연관성을 살펴보고 한우 월령과 기상 환경요인을 이용하여 한우 증체량을 예측하는 모델을 제안하고자 한다. 일반적인 통계적 모형인 가중 선형 회귀모형과 확률과정을 기반한 가우시안 프로세스 회귀모형을 통해 기상 환경요인들에 따라 한우 체중의 변화를 예측한다. 또한, 여러 개체들이 시간에 걸쳐 추적된 데이터라는 점에서 개체별로 가우시안 프로세스를 적용해야 할 필요성을 고려하여 가우시안 프로세스 패널 회귀모형을 비교 분석한다. 궁극적으로 한우 성장에 영향을 주는 중요한 환경요인을 알아내고 정밀한 사양 관리 조절이 가능하도록 하는 의사결정 시스템을 제공하고자 한다.

## 2. 분석자료 및 분석방법

### 2.1. 분석자료

연구 자료는 한우를 사육하는 전라남도 지역의 두 농가로부터 수집된 한우 생체 정보와 기상환경 자료이다. 기상환경은 기상청에서 제공하는 중관기상관측(ASOS) 자료이며, 각 관측값을 기준으로 평균(mean), 최대(max), 최소(min), 누적(cum) 등 새로운 변수를 생성하였다. 한우 데이터는 2021년 8월부터 2022년 6월까지 사육 기간 동안(26주~52주)의 주별 자료이며, 한우 개체별로 조사 기간은 약간 상이하다. 반응변수는 개체별 한우의 증체량이며, 설명변수는 월령(태어난지 몇 개월 되었는지)의 생체적 특성과 최고 기온(°C), 평균 습도(%), 최대 풍속(m/s), 평균 강수량(mm), 누적 일조(hr) 등 5가지 환경요인이다. 시차 상관분석을 통해  $t$ 시점의 한우의 생체정보는 바로 직전 ( $t-1$ ) 시점의 환경의 영향을 가장 많이 받는 것을 확인하였고, 생육 일자에 따라 환경 데이터를 주별 자료로 만들었다.

### 2.2. 분석방법

먼저, 한우의 무게와 5가지 환경정보와의 관련성을 추세 그래프를 통해 대략적으로 파악하고 상관분석을 실시하여 통계적으로 검증한다. 두 번째, 가중 선형 회귀모형은 가중 최소제곱법을 이용하여 각 데이터 포인트의 중요성을 반영하고 오차의 분산을 같게 만들어준다. 시간 변화량( $\Delta t_i$ ) 동안 한우의 증체량( $\Delta y_i$ )은 다음과 같은 선형회귀모형으로 정의된다. 여기서  $x_{ki}$ 는 주어진 기간 동안의 월령과 환경요인들이며, 오차항  $\epsilon_i$ 은 평균이 0이고, 분산이  $\Delta t_i$ 를 갖는 정규분포를 따른다.

$$\Delta y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \epsilon_i \quad (2.1)$$

따라서 시작점( $i=1$ )부터 마지막시점( $i=n$ )까지 한우의 증체량( $\Delta y$ ) 모형은 평균벡터가  $\Delta \mu$ , 공분산행렬이  $\Sigma$ 를 따르는 다변량 정규분포를 따른다. 이때, 주어진 가중 선형 회귀모형에서 회귀계수는 가중 최소제곱법을 사용하여 식 (2.2)와 같이 추정된 회귀계수를 얻을 수 있다.

$$\hat{\beta} = (X^T \Sigma X)^{-1} X^T \Delta y \quad (2.2)$$

$\Delta t_i$ 에서 한우의 증체량의 추정량은 식 (2.3)과 같이 주어지며, 시작점부터 마지막 시점까지의 한우



의 무게의 증가곡선은 식 (2.4)로 정의된다. 여기서  $w_i$ 는  $i$ 시점에서의 한우의 체중,  $y_i$ 은 한우의 초기 체중을 나타내고,  $cumsum(\Delta y_i)$ 는  $i$ 시점까지의 증가체중의 누적분을 나타낸다.

$$\Delta y_i = (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki}) \Delta t_i \quad (2.3)$$

$$w_i = y_i + cumsum(\Delta y_i), \quad i = 1, \dots, n \quad (2.4)$$

세 번째, 가우시안 프로세스는 입력 공간에서 함수의 분포를 생성하고, 이 분포를 통해 함수값이 어떤 범위 내에서 확률적으로 변동하는지를 예측한다. 데이터 분포를 구축할 뿐 아니라 예측에 대한 불확실성을 제공하여 유용한 정보를 얻을 수 있다. 또한, 커널을 이용하여 사전 정보를 통합하고 무한한 수의 함수에 대한 분포를 정의하여 회귀를 실행한다.

네 번째, 가우시안 프로세스 패널 회귀는 여러 시점에서 측정된 여러 개체의 다변량 데이터로 패널 데이터를 정의하는 방법이며, 고차원적인 관계를 포함하는 비모수적 모델을 허용한다. 각 개체별 시계열이 동일한 가우시안 프로세스를 구현하고 개체별 가우시안 프로세스가 상호 독립적이라고 가정한다. 따라서 시간이 지남에 따라 생기는 각 개체별 기울기 차이를 허용하게 된다.

### 3. 요약 및 향후 연구

본 논문은 2021년 8월부터 2022년 6월까지 주별 한우 데이터를 이용하여 한우 성장에 영향을 주는 환경요인들을 알아보았다. 한우의 무게와 월령 및 온도, 습도, 풍속, 강수량, 일조 등 5가지 환경요인들과의 관계는 상관분석을 통해 통계적으로 검증하여 유의미한 요인들을 확인하였다. 시간 변화량 동안의 한우의 증체량을 예측하기 위하여 가중 선형 회귀 모형을 사용하였고 성장 곡선에 대한 분포와 예측의 불확실성에 대한 정보를 얻기 위해 가우시안 확률과정 모형을 실시하였다. 또한, 여러 한우 개체들이 시간에 걸쳐 추적된 데이터라는 점에서 개체별로 가우시안 프로세스를 적용해야 할 필요성을 고려하여 가우시안 프로세스 패널 회귀모형을 비교 분석하였다. 본 연구는 한우 성장에 영향을 주는 중요한 환경요인을 알아내고 사양 관리를 정밀하게 조절할 수 있는 시스템을 제공하고자 진행되었다. 추후 연구에서는 음이시간, 음수시간, 음향 시그널 등에 관한 추가적인 정보가 수집되어 가축 성장에 영향을 주는 다양한 요인을 고려하여 정밀한 사양 관리 의사결정을 지원하는 시스템을 제공할 것으로 기대된다.

### References

Bernardo, J. M., Berger, J. O., Dawid, A. P., A. F. M. S. Smith. (1998). Regression and classification using Gaussian process priors. *Bayesian statistics*, 6, 475.

Karch, J. D., Brandmaier, A. M., Voelkle, M. C. (2020). Gaussian process panel modeling—machine learning inspired analysis of longitudinal panel data. *Frontiers in Psychology*, 11, 351

Kiers, H. A. (1997). Weighted least squares fitting using ordinary least squares algorithms. *Psychometrika*, 62, 251-266

MacKay, D. J. (1998). Introduction to Gaussian processes. *NATO ASI series F computer and systems sciences*, 168, 133-166

Na, M. H., Park, Y. H., Cho, W. H. (2017). A study on optimal environmental factors of tomato using smart farm data, *Journal of the Korean Data & Information Science Society*, 28(6), 1216-1226

Wang, J. (2020). An intuitive tutorial to Gaussian processes regression. arXiv preprint arXiv:2009.10862

## 감성분석에서의 BERT와 GPT 모형 비교 - D대학교 서술형 강의평가 분석

성진용<sup>1</sup>, 권민수<sup>2</sup>, 심성현<sup>3</sup>, 최승배<sup>4</sup>, 강창원<sup>5</sup>

### 요약

최근 생성형 인공지능(Generative AI) 알고리즘이 많은 분야에서 활용되고 있는데 특히 텍스트 분석 및 자연어 처리 문제에서 트랜스포머 계열의 딥러닝 모형이 좋은 성능을 보이고 있다. 트랜스포머 계열의 딥러닝 언어 모형은 일반적으로 사전학습(pre-training)과 파인튜닝(fine-tuning) 두 부분의 과정으로 이루어져 있으며 대표적 모형으로 BERT(Bidirectional Encoder Representations from Transformers)와 OpenAI GPT(Generative Pre-trained Transformers)를 들 수 있다. 본 연구에서는 텍스트 마이닝에서 자주 등장하는 감성분석을 BERT와 GPT모형을 적용하여 비교하고자 한다. 이를 위해 첫째, BERT와 GPT 알고리즘의 차이를 설명하고 둘째, D대학교 서술형 강의평가 자료에 대하여 긍·부정 감성분석을 하였다. 셋째, 적용한 두 모형의 오분류 데이터를 이용하여 두 모형 간 오류 특성을 비교하였다. 한편 분석을 위하여 강의평가 자료(총 10,061 문장)를 학습용 데이터 70%, 검증용 데이터 30%로 분할하였다. BERT 모형은 사전학습 과정과 파인튜닝 과정(학습용 데이터 70%)을 거친 후 검증용 데이터셋(30%)에 적용하였으며, 반면 GPT 모형은 버전 GPT-3.5를 이용하였는데 알고리즘 특성 상 사전학습 과정만을 거친 후 동일한 검증용 데이터셋에 적용하였다. 분석 결과, 검증용 데이터에 대한 예측력 측면에서 BERT 모형의 정분류율이 94.6%(민감도 97%, 특이도 86%)로 나타나 정분류율 88%를 보인 GPT(민감도 95.4%, 특이도 48%)방법에 비해 우수한 성능을 보였다.

주요용어 : 텍스트 마이닝, 트랜스포머, 감성분석, BERT, GPT

<sup>1</sup>47340 부산광역시 부산진구 가야동 산 24 동의대학교 대학원 수확정보통계학과, 석사과정.

E-mail: wlsdyd3194@naver.com

<sup>2</sup>48059 부산광역시 해운대구 센텀북대로 60 ㈜에스티이노베이션, 대표이사. E-mail: ceo@stinnovation.co.kr

<sup>3</sup>47340 부산광역시 부산진구 가야동 산 24 동의대학교 산업경영·빅데이터공학전공, 교수.

E-mail: ssh@deu.ac.kr

<sup>4</sup>47340 부산광역시 부산진구 가야동 산 24 동의대학교 산업경영·빅데이터공학전공, 교수.

E-mail: csb4851@deu.ac.kr

<sup>5</sup>47340 부산광역시 부산진구 가야동 산 24 동의대학교 산업경영·빅데이터공학전공, 교수.

E-mail: cwkwang@deu.ac.kr

## 딥러닝 기반 시계열 분석 모형의 불확실성 정량화 비교 연구\*

윤영인<sup>1</sup>, 정혜영<sup>2</sup>

### 요약

인공지능의 발전으로 머신러닝과 딥러닝 모형이 성능이 좋아 다양한 산업에서 적용되고 있고 최근 금융시장에서도 적용되는 사례가 증가하고 있다. 금융시장의 시계열 데이터의 경우 여러 가정을 만족해야하는 전통적인 통계 모형과 달리 딥러닝 모형은 비교적 모형 설정이 간단하며 성능이 좋아 딥러닝 모형을 활용하는 논문이 증가추세를 보이고 있다. 그러나 딥러닝 모형은 예측 결과가 나오게 된 과정과 해석을 파악하기에 어려움이 있어 결과에 대한 해석이 특히 중요시 되는 금융에 딥러닝 모형을 적용하는데 장애 요소가 되고 있고 신뢰할 수 있는 모형에 대한 필요성이 대두되고 있다. 신뢰할 수 있는 모형이란 모형의 변동에도 일관된 예측을 보이는 안정적인 모형로서 모형의 불확실성을 통해 파악할 수 있다. 본 연구의 사전 연구로 전통적인 통계모형 ARIMA(AutoRegressive Integrated Moving Average)보다 여러 딥러닝 모형에서 성능이 좋음을 확인하였고 다양한 패턴의 시계열 데이터에서 LSTM(Long Short Term Memory) 모형의 불확실성을 측정하는 방법에 대한 비교 연구를 진행하였다. 본 논문은 다양한 패턴의 시계열 데이터에 대해 딥러닝 모형의 불확실성을 확인하고 이상치를 탐지하는지 파악하여 신뢰할 수 있는 모형의 기준을 보이고자 한다. 딥러닝 모형은 시계열 데이터에 주로 쓰이는 CNN(Convolutional Neural Network), LSTM, MLP(Multi Layer Perceptron), 및 CNN-LSTM 모형을 적용하였고 MC Dropout을 통해 베이지안 관점에서 불확실성을 측정하였다. 실험 결과 불확실성은 성능이 유사한 모형간의 차이가 있으며 성능이 가장 우수하지는 않아도 불확실성이 적어 안정적인 모형이 존재함을 확인하였다. 이를 통해 불확실성이 모형의 정확도와 함께 모델 선택 시 고려되어야 할 요소임을 제안하고자 한다.

주요용어 : 딥러닝, 시계열, 안정적인 모형, 불확실성, MC Dropout

\* 본 논문은 한국연구재단 연구자지원사업(NRF-2022R1F1A1074939)의 지원을 받아 수행된 연구임.

<sup>1</sup>15588 안산시 상록구 한양대학교 55, 한양대학교 일반대학원 응용수학과, 석사과정.

E-mail: yyi99@hanyang.ac.kr

<sup>2</sup>15588 안산시 상록구 한양대학교 55, 한양대학교 ERICA 수리데이터사이언스학과, 부교수.

E-mail: hyjunglove@hanyang.ac.kr

## 중국의 호주산 석탄 수입 금지가 원료탄 가격에 미친 영향: 반사실적 분석을 중심으로

김민수<sup>1</sup>, 임석<sup>2</sup>, 유지애<sup>3</sup>

### 요약

2020년까지만 해도 중국은 매년 7,000~8,000만 톤의 석탄을 호주로부터 수입하였다. 하지만 중국과 호주 사이의 정치적 갈등이 고조되던 2020년 10월 중국이 호주산 석탄통관을 중단했고, 2022년 중국의 호주산 석탄 수입량은 2020년 대비 4%에 못 미쳤다. 이후 호주의 정권 교체와 함께 양국간 관계가 개선된 후 2023년 1월 중국은 호주산 석탄 수입을 정식으로 재개하기로 결정하였다. 본 연구는 중국의 호주산 석탄 수입 금지 정책이 호주산 제철용 원료탄 가격에 미친 인과적 효과(Causal Effect)가 시점마다 어떻게 나타나는지 분석하기 위해 반사실적 분석(Counterfactual Analysis)을 진행하였다. 반사실적 분석을 위해 BSTS (Bayesian Structural Time Series) 모델을 사용하였으며, 종속변수로는 제철용 원료탄 동호주 FOB 가격을, 설명변수로는 정책에 영향을 받지 않는 인도 조강 생산량과 달러 인덱스를 사용하였다. 분석 결과 정책의 효과는 8~10개월 후에 나타나기 시작한 것으로 확인되었으며, 정책의 시행으로 인해 원료탄 가격이 톤당 200~300 달러 더 상승하였음을 파악할 수 있었다. 한편 호주산 제철용 원료탄 수입 비중이 60%에 달하는 한국의 입장에서 중국의 호주산 제철용 원료탄 수입 재개에 따른 가격 변화에 민감할 수 밖에 없다. 정치적 이해관계에 따른 급작스러운 수출입 정책 변화가 발생할 수 있는 만큼 이와 같은 케이스 분석은 국가 및 기업 측면에서 원자재 확보를 위한 노력에 도움이 될 것으로 판단된다.

주요용어 : 인과추론, counterfactual, BSTS, 제철용 원료탄, 정책효과.

<sup>1</sup>04034 서울시 마포구 양화로 93, (주)코리아피디에스 연구원, Email: key1036@koreapds.com

<sup>2</sup>(교신저자) 04034 서울시 마포구 양화로 93, (주)코리아피디에스 책임연구원, Email: seoklim@koreapds.com

<sup>3</sup>04034 서울시 마포구 양화로 93, (주)코리아피디에스 연구원, Email: yeriyou@koreapds.com

## 다중 스펙트럼 영상과 Video Vision Transformer를 이용한 벼의 수확량 예측\*

김다현<sup>1</sup>, 나명환<sup>2</sup>

### 요 약

정부와 농민단체는 매년 쌀을 얼마나 생산할 수 있을지 예측하는 문제에 많은 관심을 기울이고 있다. 벼의 수확량을 예측하는 것은 정부에서 쌀의 수급을 적절히 조절할 수 있는 정책을 수립하는 데 주요한 요인일 뿐만 아니라 농산물가격 안정과 농가의 재배 결정에도 영향을 미치게 된다. 하지만 해마다 변하는 기상요인들과 다양한 병충해와 같은 변동요인들로 인하여 벼의 수확량을 정확하게 예측하는 것은 어려운 문제이다. 본 연구에서는 무인항공기에 탑재된 다중 스펙트럼 센서를 통해 벼의 생육기간 동안 이미지를 여러 번 수집하고, 딥러닝 알고리즘을 이용하여 벼 수확량을 예측하였다. 다중 스펙트럼 이미지는 일정 기간을 두고 여러 번 촬영된 일종의 영상 데이터로 볼 수 있으며, 딥러닝에서 Transformer 구조를 영상 컴퓨터 비전에 적용한 Video Vision Transformer(ViViT) 모델을 사용해 벼 수확량을 예측하였다. ViViT 모델은 입력 영상을 일정한 크기로 분할한 patch들을 생성하는데, 이 patch의 크기를 다르게 설정하여 모델을 학습한 결과 작은 patch 크기를 사용할수록 예측력이 좋아지는 것으로 나타났다. 또한 이미지 처리 분야에서 사용되어온 CNN 구조에서 영상을 입력으로 받는 3D CNN 모델과 예측 성능을 비교해본 결과 작은 patch 크기를 사용한 ViViT 모델의 성능이 더 우수한 것으로 나타났다. ViViT 모델의 가중치 Heat map을 시각화한 결과 8월 중후반에 촬영된 이미지가 중요하게 나타나 벼를 수확하기 약 두 달 전에 수확량 예측이 가능할 것으로 보인다.

\*본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원의 노지분야 스마트농업기술 단기 고도화 사업의 지원을 받아 연구되었음(32204003).

<sup>1</sup>(제1저자) 61186, 광주광역시 북구 용봉로 77, 전남대학교 수학과/통계학과 석사과정. Email: qhqk7132@jnu.ac.kr

<sup>2</sup>(교신저자) 61186 광주광역시 북구 용봉로 77, 전남대학교 통계학과 교수. Email: nmh@chonnam.ac.kr

## 국민건강보험빅데이터를 이용한 한국의 세대 간 소득이동성 연구

전세봄<sup>1</sup>, 김준일<sup>2</sup>

### 요 약

세대 간 갈등 및 불평등 문제는 한국 뿐 아니라 전 세계적으로 대두되고 있으며, 특히 결과에 대한 불평등 뿐 아니라 결과에 이르는 과정 및 기회의 불평등에 대한 비판론이 확산되고 있다. 본 연구는 국민건강보험빅데이터 부모와 자녀의 소득분위를 이용하여 한국 사회의 기회불평등 및 세대 간 소득이동성을 분석함으로써 부모세대와 자녀세대의 소득 간의 관계를 파악하였다. 또한 자녀세대 성과에 영향을 미치는 요인을 파악하기 위하여 계층이동 사다리의 역할을 하는 교육이 거주지역에 따른 환경 및 교육의 질에 의존하므로 이러한 환경요소를 반영하여, 소득 이동성 같은 기회불평등이 교육기회나 주거환경, 지역 등에 어느 정도 영향을 받는지 알아보려고 하였다. 아버지 소득분위와 자녀의 소득분위는 연관성이 있었으나, 소득이 일정 수준 이상에서만 아버지 소득에 따라 자녀 소득이 달라지는 것으로 나타났으며 그 외에는 기회 불평등이 두드러지니 않는 것으로 나타났다.

주요용어 : 기회불평등, 소득분위, 로그선형모형, 순서형분할표, 조건부분포.

---

<sup>1</sup>35349 대전광역시 도안북로 88 목원대학교 사회과학대학 마케팅빅데이터학과 부교수.  
E-mail: alwaysns@mokwon.ac.kr

<sup>2</sup>(교신저자) 35349 대전광역시 도안북로 88 목원대학교 사회과학대학 금융경제학과 부교수.  
E-mail: kjoonil@mokwon.ac.kr

## A Study on Efficient Data Synthesis Using Deep Generative Models

*Jiwoo Kim<sup>1</sup>, Seri Park<sup>2</sup>, Dongha Kim<sup>3</sup>*

### Abstract

As data sharing has received much attention, there is a growing concern about the risk of privacy exposure. To overcome the trade-off between data sharing and privacy risk, generating and distributing synthetic data, instead of original data, is considered as a good alternative. Synthetic data refer to simulated data of the original ones, but still containing meaningful information. Well-generated synthetic data can encourage active data sharing while protecting the privacy of the original data. In this study, we propose a new deep generative method to synthesize data mixed with categorical and continuous variables. Our method is based on the likelihood approach and focuses on effectively handling data with imbalanced and complex distributions. We adopt a conditional generative model to synthesize imbalanced categorical data successfully. As for the continuous case, we devise a learnable pre-processing technique that changes a given distribution to a simple and easy-to-generate distribution. We conduct various experimental analyses to demonstrate that our method is excellent and efficient in generating synthetic data compared to other recent competitors.

*Keywords* : synthetic data generation, deep generative models, importance weighted autoencoders.

---

<sup>1</sup>Graduate Student, Department of Statistics, Sungshin Women's University, 2, 34 da-gil, Bomun-ro, Seongbuk-gu, Seoul, 02844, Korea. E-mail: kjiwoo6870@gmail.com

<sup>2</sup>Graduate Student, Department of Statistics, Sungshin Women's University, 2, 34 da-gil, Bomun-ro, Seongbuk-gu, Seoul, 02844, Korea. E-mail: seri.jpark96@gmail.com

<sup>3</sup>(Corresponding Author) Assistant Professor, Department of Statistics and Data Science Center, Sungshin Women's University, 2, 34 da-gil, Bomun-ro, Seongbuk-gu, Seoul, 02844, Korea.  
E-mail: dongha0718@sungshin.ac.kr

## YOLO와 XAI를 활용한 궤양병 분류 알고리즘 연구

장진결<sup>1</sup>, 김기환<sup>2</sup>

### 요약

인공지능(Artificial Intelligence, AI)은 컴퓨터가 학습하고 추론하는 기술로, 인간의 지적 능력과 유사하게 작동한다. 인공지능은 빅데이터와 머신러닝 기술의 발전으로 다양한 분야에서 혁신을 가져왔으며, 전 세계적으로 많은 관심을 받고 있으며, 다양한 국가에서 인공지능에 대한 전략을 수립하여 발전을 추진하고 있다. 그러나 이러한 인공지능은 입력과 출력 사이에 어떠한 내부 과정이 발생하는지 알 수 없는 블랙박스 모델로서 모델이 도출한 결과에 대한 설명력이 부족한 한계를 가지고 있다. 이를 해결하기 위해 설명 가능한 인공지능(Explainable AI, XAI) 기술이 개발되고 발전하고 있다. 인공지능은 다양한 산업 분야에서 활용되고 있으며, 스마트팜과 같이 농업에서도 생산성 향상을 위해 사용된다. 본 연구에서는 감귤의 대표적인 질병인 궤양병의 발생 여부를 분류하기 위해 정상군과 궤양병 감귤 및 잎 이미지 데이터를 확보하였으며 딥러닝을 기반으로 궤양병 분류 알고리즘을 구축하였다. 확보한 데이터는 감귤과 잎을 개별 촬영한 이미지로 실생활에서의 활용이 어렵다. 따라서 실제 감귤 나무 이미지를 수집하여 감귤 나무 내 감귤의 궤양병 여부를 분류하였다. 이 과정에서 대표적인 객체 검출 알고리즘인 YOLO(You Look Only Once)를 활용하여 감귤을 검출하였으며, XAI를 활용하여 분석 결과를 점검하였다.

주 용어 : 인공지능, 객체 검출, XAI, YOLO, 궤양병.

---

<sup>1</sup>30019 세종특별자치시 세종로 2511, 고려대학교 일반대학원 경제통계학과 국가통계전공 석사과정.  
E-mail: torres5895@korea.ac.kr

<sup>2</sup>(교신저자) 30019 세종특별자치시 세종로 2511, 고려대학교 경제통계학부 국가통계전공 교수.  
E-mail: korpen@korea.ac.kr



## Correlation-Based Clustering Analysis of the Opioid Epidemic Data on East Texas Counties

*Sahar Soliman<sup>1</sup>, Doo Young Kim<sup>2</sup>, Noura Ammar<sup>3</sup>, Jocelynn Robinson<sup>2</sup>, Melinda Holt<sup>2</sup>, Kevin Lord<sup>1</sup>, Maria Zimpfer<sup>4</sup>, David Hoffpauir<sup>4</sup>, Diego Alvarez<sup>1</sup>*

### Abstract

For the past several decades, the United States has been grappling with an opioid epidemic that has disproportionately impacted vulnerable and disadvantaged communities. Initially, the crisis was believed to have been caused by over-prescription of opioids and aggressive pharmaceutical marketing tactics. However, more recent evidence suggests that the rise of the opioid crisis is also tied to increased community vulnerability brought about by economic and environmental factors. This study aims to determine the association between county-level demographic and socioeconomic characteristics and prescription opioid use in predominantly rural East Texas. Demographic and socioeconomic characteristics of 35 counties comprising the Public Health Region 4/5 North (PHR 4/5N) of Texas were analyzed using the correlation-based cluster analysis. Prescription Drug Monitoring Program (PDMP) opioid utilization data was correlated with socioeconomic and socio-determinants of health within the counties in the same cluster. Texas has been considered one of the lowest opioid-prescribing states while East Texas has an average opioid prescription rate exceeding the national average. In this talk, I will mainly present how the correlation-based clustering analysis was performed with some basics of clustering analysis, and the significance of the interpretation of our clustering results will be discussed.

---

<sup>1</sup>Department of Physiology and Pharmacology, College of Osteopathic Medicine, Sam Houston State University, Conroe, Texas.

<sup>2</sup>Department of Mathematics and Statistics, College of Science and Engineering Technology, Sam Houston State University, Huntsville, Texas.

<sup>3</sup>Department of Public Health and Preventive Medicine, College of Medicine, Ain Shams University, Cairo, Egypt.

<sup>4</sup>Department of Mathematics and Statistics, College of Science and Engineering Technology, Sam Houston State University, Huntsville, Texas.

## Improving the Usability of Collaborative Robot based on Data Analysis<sup>\*</sup>

*Jeong Soyeong*<sup>1</sup>, *Christoffer Sloth*<sup>2</sup>

### Abstract

The demand for collaborative robots(cobots) is growing steadily. However, they are still not easily visible in our surroundings and their usage remains limited. One reason cobots are difficult to use is their blunt movements. They are difficult for humans to move easily and to respond quickly when encountering obstacles. This study is to utilize the data to make the cobot move with less power. A spring-damper model is used to describe the human arm that moves the cobot. However, it is difficult to obtain accurate data due to the limitations of sensing. At the endpoints, stiffness and damping should be estimated. The stiffness at the endpoint will vary because the speed and posture change whenever a person moves the robot. This is a characteristic of kinaesthetic learning. In this study, the human behavior was predicted based on the constrained minimum jerk model using the data of a person's movement. Based on the prediction, we make robots help the person and we show robots allow humans to move robots with less force.

*Keywords* : collaborative robots, human motion prediction, arm compliance estimation, kinaesthetic demonstrations.

---

<sup>\*</sup>This research was supported by the MOTIE (Ministry of Trade, Industry, and Energy) in Korea, under A Project to foster global talent for innovation in the robot industry (P0017311) supervised by the Korea Institute for Advancement of Technology (KIAT).

<sup>1</sup>(Corresponding Author)Guest researcher, Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark. E-mail: soje@sdu.dk, soyeong23@gmail.com

<sup>2</sup>Associate Professor, Maersk Mc-Kinney Moller Institute, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark. E-mail: chsl@sdu.dk

## 기업평가를 위한 기업정보의 다각적 활용 사례

이연경<sup>1</sup>, 김종운<sup>2</sup>

### 요약

기업정보는 기업의 생애주기 동안 기록 및 획득할 수 있는 모든 정보를 의미하며, 기업정보에는 기업의 경영 성과, 차입금에 대한 상환능력 및 현금지급능력, 자산의 효율적인 운영능력 등을 측정할 수 있는 재무정보가 있으며, 재무정보 외에 기업이 속한 산업 전망, 경영관리 능력 및 경영 안전성, 영업자원 및 구매<생산>판매위험, 질적재무위험, 거래 신뢰도 및 업체 신뢰도 등 거래활동을 평가할 수 있는 비재무적 정보가 있다. 본 발표에서는 특정 시점에 수집할 수 있는 다양한 기업정보를 소개하고, 이를 활용하여 최근 금융기관 및 공공기관에서 시도하고 있는 기업 평가 사례를 알아보고자 한다.

주요용어 : 기업정보, 기업평가, 금융기관, 공공기관.

---

<sup>1</sup>NICE평가정보. E-mail: yklee@nice.co.kr

<sup>2</sup>NICE평가정보.

## 제품 불량 검출을 위한 딥러닝 모델 개발 사례

김민정<sup>1</sup>

### 요약

제조업 분야에서는 제품 생산 후 품질을 확인하기 위해 주로 작업자의 육안을 통해 외관검사를 수행해왔다. 하지만 작업자의 주관 및 피로도에 따라 불량 검출 기준이 달라질 수 있고 운영 비용이 많이 든다는 단점으로 인해, 딥러닝 모델 기반의 자동화 검사 시스템을 도입하려는 움직임이 늘고 있다. 불량을 검출하기 위한 딥러닝 모델은 대부분 이미지 기반의 컴퓨터 비전(Computer Vision) 분야 알고리즘을 사용하여 학습한다. 이를 위해서는 제품 외관 이미지 수집이 선행되어야 하며, 수행 목적과 운영 환경에 맞는 분석 기법 선택이 필요하다. 본 발표에서는 컴퓨터 비전의 객체 탐지(Object Detection) 기법을 사용한 모델 개발 사례를 소개하고자 하며, 그 과정에서 모델의 핵심 알고리즘 구조와 이미지 데이터 분석에서 가장 중요한 단계 중 하나인 라벨링(Labeling)에 대해서도 알아보하고자 한다.

주요용어 : 딥러닝, 객체 탐지 모델, 이미지 분석, 라벨링, 제조업.

---

<sup>1</sup>데이터솔루션 데이터사이언스본부 책임연구원. E-mail: mjkim@datasolution.kr

## 사회안전지수와 지역주민 인식에 대한 분석

박현수<sup>1</sup>, 장안식<sup>2</sup>

### 요약

본 연구는 '사회안전지수'를 통해 시민의 불안감에 영향을 미치는 개인적 요인과 지역적 요인을 살펴보았다. 분석방법으로 위계적 선형 모형을 사용하였는데, 독립변수로 거주지역에 대한 주관적 인식과 객관적 자료(정량지표)를 사용하였다. 미시적 수준의 개인적/주관적 인식과 거시적 수준의 객관적 정량지표가 시민의 불안감(종속변수)에 미치는 영향을 동시에 살펴보기 위해서 '위계적 선형 모형(Hierarchical Linear Model, HLM)을 사용하여 분석하였다. 분석은 R의 'lme4' 패키지를 사용하였다. 개인수준에서는 '건강보전 만족도'를 제외한 '경제활동 만족도', '생활안전 만족도', '주거환경 만족도'가 시민의 불안감에 유의한 영향을 미쳤고, 지역수준에서는 '주거환경 정량지표'만이 유의한 영향을 미치는 것으로 나타났다. 지역의 경제활동, 생활안전, 건강보전에 관한 객관적 지표들은 지역 주민이 가시적으로 잘 인식하기 어려우나, 이에 비해 주거환경과 관련된 객관적 지표들은 상대적으로 지역 주민이 체감할 수 있어 개인의 불안감에 영향을 미치는 것으로 볼 수 있다. 마지막으로, 연구결과를 바탕으로 시민의 불안감을 줄일 수 있는 정책적 대안에 대해서 논의 하였다.

주요용어 : 불안, 사회안전지수, 위계적 선형 모형(Hierarchical Linear Model: HLM).

---

<sup>1</sup>28644 충청북도 청주시 서원구 흥대로 1, 충북대학교 국가위기관리연구소 시민치안연구센터장.  
E-mail: cha0s51@hanmai.net

<sup>2</sup>(교신저자) 06650 서울특별시 서초구 반포대로24길 76, 케이스탯컨설팅 공공사회정책연구소장.  
E-mail: ansik@kstat.co.kr

## IPA 활용 지역수출유망단지 진단평가\*

홍우진<sup>1</sup>, 손황배<sup>2</sup>, 서현택<sup>3</sup>, 박정민<sup>4</sup>

### 요약

본 연구는 IPA를 활용하여 2018년부터 발굴하여 종합적인 맞춤형 기술·컨설팅을 지원하고 있는 지역수출유망단지의 종합적인 역량을 진단하고 분석 결과를 토대로 정책의 대응방향을 제안하는 데 목적이 있다. 20개 단지를 대상으로 대표와 담당 지도사 총 40명에게 구조화된 진단평가표를 활용하여 조사를 진행하였다. 2사분면(중요도가 높으나 성취도가 낮은 항목)에는 물량확보 등 6개의 속성이 도출되었고, 1사분면에는 품질 매뉴얼 등 9개의 속성, 4사분면에는 외부전문가 등 11개의 속성, 3사분면에는 시장 교섭력 등 6개의 속성이 도출되었다. 이를 통하여 생산·상품화·마케팅·조직운영 등 역량별 강화를 위한 교육 및 컨설팅 맞춤형 프로그램 설계에 활용도를 높여갈 수 있을 것이다. 향후 연구에서는 지역수출유망단지 특성별로 상이하게 나타나는 결과들에 대하여 차이점을 비교하고, 역량수준과 운영현황(수출액 등)의 영향관계를 분석하여 종합적으로 보완하거나 영역별·항목별 상대적인 중요도를 감안해 가중치를 부여하는 방법을 고려해볼 수 있다.

주요용어 : IPA, 신선농산물, 지역수출유망단지, 역량, 진단.

### 1. 연구목적

농촌진흥청은 농산물 수출 활성화와 경쟁력 확보를 위해 2018년부터 ‘지역수출유망단지’를 발굴하여 종합적인 맞춤형 기술·컨설팅을 지원하는 정책과제를 추진해오고 있으며, 2022년까지 30개 품목을 목표로 수출유망품목을 발굴해 지원할 계획이며, 2022년 현재 20개 품목에 대하여 지역수출유망단지(총 20개소)를 선정하여 집중 지원·관리 중이다.

지역수출유망단지의 수출역량을 강화하기 위해서는 현재의 역량수준을 보다 객관적으로 평가하여 진단할 수 있는 지표를 개발하고 이를 활용한 조사 및 분석 결과를 토대로 정책의 대응방향을 제안하는 데에 주요한 목적이 있다.

\*본 논문은 농촌진흥청의 지원을 받아 연구된 것임(PJ015694).

<sup>1</sup>58213 대한민국 전라남도 나주시 산포면 세남로 1508, 전라남도농업기술원 농업연구사.

E-mail: wojin92@korea.kr

<sup>2</sup>54875 대한민국 전라북도 전주시 덕진구 농생명로 300, 농촌진흥청 농업연구사. E-mail: misson7@rda.go.kr

<sup>3</sup>24203 대한민국 강원특별자치도 춘천시 신북읍 장본길 40, 강원특별자치도농업기술원 농업연구사.

E-mail: gusxor0000@korea.kr

<sup>4</sup>41404 대한민국 대구광역시 북구 칠곡중앙대로136길 47, 경상북도농업기술원 농업연구사.

E-mail: pjmgangela@korea.kr

## 2. 이론적 배경

Kim(2017)은 농촌진흥청 표준진단표 등 농업법인에 대한 기존 경영분석 방법의 한계를 지적하고, 국내외 농업경영체 경영분석 진단방법을 비교(대상·구성·특징·장단점 등) 분석하여 경영진단 지표를 개발, 이를 토대로 농업법인의 품목 유형과 특성을 반영한 BSC관점의 재무적·비재무적 진단 지표를 발표하였다.

Hwang 외(2017)는 품목농업인연구회의 성격진단을 통해 운영·학습·사업·연계·지원 역량의 관점에서 평가지표를 발굴하여 IPA분석, 조직역량과 성과효과 관계의 역량별 매개효과를 분석하여 품목농업인연구회의 산지조직화 실천력을 높이고, 농촌진흥사업의 효율성을 제고하기 위한 운영 활성화 방안으로 역량강화 및 지원강화 프로그램 등을 제안하였다.

Kim, Han(2015)은 수출업체 120개(선도 16, 일반 104 / 과실 45, 과채 33, 화훼 23, 버섯 19)와 수출농가 550명(원예전문단지 95, 개별농가 455)을 대상으로 설문조사를 진행, 수출단계별 활동수준·운영주체·운영단계 등 3개 범주로 유형을 분류하여 운영효율성(DEA분석 투입·산출지향 모형), 운영성과 영향요인(SUR모형) 등을 분석하여 농식품수출조직의 운영실태와 문제점, 운영성과 영향요인 분석을 통해 자생력·지속성을 가진 수출조직으로 육성하기 위한 선도조직·개별업체·수출농가의 결속 성장 및 정착 등 지원방안을 제안하였다.

IPA 분석은 중요도를 Y축으로, 만족도를 X축으로 한 2차원의 매트릭스에 표시하는데, 중요도와 만족도 각 속성의 평균값을 적용한 중심점을 기준으로 4사분면에 영역을 구분하며, 각 속성들을 위치시켜 의미를 부여하고 그에 따른 시사점을 도출할 수 있다(Hwang, 2017).

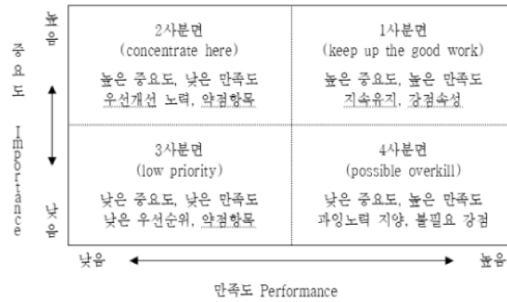


Figure 1. 중요도-만족도(IPA) 분석의 틀(Martilla & James, 1977)

## 3. 연구방법

본 연구는 지역수출유망단지의 종합적인 역량을 진단하기 위해 구조화된 진단평가지표를 활용하여 지역수출유망단지(20개소) 대표와 담당 지도사 총 40명을 대상으로 2022년 7월 18일부터 8월 8일까지 조사를 실시하였다.

설문은 전반적인 운영현황과 기대효과/필요사항, 역량수준, 역량진단을 평가하는 질문으로 구성되었으며 운영현황은 ①단지현황, ②고용인력, ③수출 및 유통 현황, ④상품화 및 시설기반, ⑤인증 및 자조금, ⑥기타 등 전체 6개 영역을 구분하여 평가항목을 설계하였다.

역량진단에서는 종합적인 역량진단을 위한 4개 영역(①생산역량, ②상품화역량, ③수출마케팅역량, ④조직운영역량)을 설정하고, 전체 32개의 세부적인 평가항목을 구성하였다.

지역수출유망단지의 상황을 고려해 평가항목별에 대하여 중요하게 생각하는 정도(중요도)와 현재의 역량수준(성취도)을 동시에 평가하도록 구성하여 IPA 분석에 활용하고 평가방식은 리코드 7점 척도를 활용하여 각각의 세부 평가항목에 대하여 인식하는 정도를 응답하도록 하였다.

### 4. 분석결과

지역수출유망단지의 운영형태는 영농조합법인이 가장 많은 11개소로 55.0%를 차지하였고, 이어서 지역농협(20.0%), 작목반(15.0%), 기타(10.0%) 등 순이고 전문단지 지정현황은 농산물전문생산단지(aT센터)가 11개소 지정되었으며 전체의 55.0% 차지, 수출통합조직 2개소, 수출선도조직 2개소 등으로 지정되어 있다.

Table 1. 수정 IPA 분석에 따른 지역수출유망단지 속성별 분석값

no	세부항목	명시적 성취도	내재적 중요도	no	세부항목	명시적 성취도	내재적 중요도
1	생산기반	5.00	0.21	19	정보수집	4.28	-0.37
2	생산매뉴얼	4.93	0.50	20	시장교섭력	4.60	0.06
3	물량확보	4.88	0.11	21	가격변동	3.55	0.29
4	적합품종	4.73	-0.37	22	업체관계	5.08	0.44
5	표준화	5.43	-0.09	23	클레임 등	5.30	-0.08
6	안전성	5.58	-0.35	24	운영규정	5.28	0.28
7	기록관리	5.58	-0.10	25	의사체계	5.23	-0.22
8	계약재매	5.10	0.40	26	운영계획	5.63	-0.09
9	유통물류	4.98	0.45	27	교육컨설팅	5.73	0.39
10	관리지장	4.75	0.43	28	경영비절감	5.53	0.05
11	공동작업	5.23	-0.13	29	공동자금	5.28	0.20
12	품질매뉴얼	5.33	0.47	30	연계협력	5.28	0.59
13	전문인력	5.00	-0.01	31	행정연계	5.33	-0.47
14	상품화	4.50	0.18	32	외부전문가	5.18	-0.03
15	이력추적	4.65	0.00				
16	마케팅전략	4.13	-0.15		<b>평균</b>	<b>4.97</b>	<b>0.09</b>
17	전문인력	3.74	-0.17		<b>최소값</b>	<b>3.55</b>	<b>-0.47</b>
18	홍보활동	4.40	0.45		<b>최대값</b>	<b>5.73</b>	<b>0.59</b>

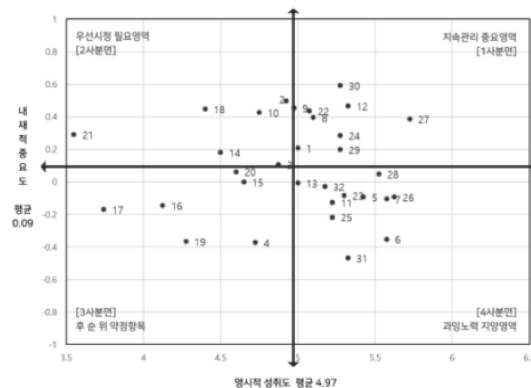


Figure 3. 수정된 중요도-성취도 분석결과



편상관계수를 활용한 수정 IPA 방법을 활용하여 직접 중요도를 간접 중요도로 변환, 속성별 값을 다음과 같이 산출하였다.

2사분면(중요도가 높으나 성취도가 낮은 항목)은 운영 활성화를 위한 우선과제에 해당되며 생산역량의 생산매뉴얼·물량확보가, 상품화역량의 관리저장·상품화가, 마케팅역량의 홍보활동·가격변동 등이 해당된다.

1사분면(중요도가 높으며 성취도도 높은 항목) 영역은 지속관리가 중요하며 생산역량의 생산기반·계약제배, 상품화역량의 유통물류·품질매뉴얼, 마케팅역량의 업체관계, 조직운영역량의 운영규정·교육 컨설팅·공동자금·연계협력 등이 해당된다.

4사분면(중요도는 높으나 성취도가 높은 항목)으로 과잉노력을 지양해야하며 생산역량의 표준화·안전성·기록관리, 상품화역량의 공동작업·품질전문인력, 마케팅역량의 검역 및 통관, 조직운영역량의 의사결정체계·운영계획·경영비절감·행정연계·외부전문가 등이 영역에 해당한다.

마지막 3사분면(중요도가 낮으며 성취도도 낮은 항목)은 후순위 대응이 적절하며 생산역량의 적합중, 마케팅역량의 이력추적·마케팅전략·수출전담인력·정보수집·시장교섭력 등이 영역에 해당한다.

## 5. 결론

본 연구는 현재 운영중인 지역수출유망단지를 IPA를 활용하여 역량을 진단하고 나아가야 할 방향을 제시하고자 한다. 우선시정 필요영역을 단기 대응과제로 설정하여 추진하고, 연차별 목표관리 계획을 수립하여 보다 체계적인 지원방식으로 전환하고 이를 통해 생산·상품화·마케팅·조직운영 등 역량별 강화를 위한 교육 및 컨설팅 맞춤형 프로그램 설계에 활용도를 높여갈 수 있을 것으로 기대된다.

연구대상인 지역수출유망단지 특성별로 상이하게 나타나는 결과들에 대하여 차이점을 비교하고 역량수준과 운영현황(수출액 등)의 영향관계를 분석하여 종합적으로 보완해갈 필요가 있고 조직역량에 대한 영역별·항목별 상대적인 중요도를 감안하여 분석결과를 고도화하기 위해서는 가중치를 부여하는 방법을 고려해볼 수 있다.

## Reference

- Gouk. S. Y. (2008). A Study on Success Factors of Farmers' Marketing Organizations. (in Korean).
- Kang. C. W. (2009). A Study on a Plan to Increase produce and agricultural foods export, *Journal of Distribution Science*, 7(4), 27-36. (in Korean).
- Kim. Y. J., Kim. W. N., LEE. M. J. (2014). An Analysis on the Impact of the FTA Support Center on Small- and Medium-sized Export Companies, *Journal of the Korean Data Analysis Society*, 16(3), 1373-1382. (in Korean).
- Roh. Y. H., Jo. L. H. (2007). An Importance-Performance Analysis of Attractiveness Attributes Assessment for Rural Tourism, *Journal of the Korean Data Analysis Society*, 9(1), 175-187. (in Korean).
- Song. N. E., Kim. B. Y., Yang. H. S. (2014). The Analysis of an Influence of the Korean Wave on Korean Processed Foods Exports, *Journal of the Korean Data Analysis Society*, 16(4), 2081-2090. (in Korean).
- Song. Y., Choi. W. H., Jun. J. K. (2020). A Study on Activation Strategy for Gamcheon Culture Village Tourism of Chinese Tourists through Importance-Performance Analysis, *Journal of the Korean Data Analysis Society*, 22(4), 1579-1592. (in Korean).



## 서울 아파트 가격에 영향을 주는 요인들

심성훈<sup>1</sup>

### 요약

본 연구는 2008년 세계금융위기 이후부터 최근까지 서울지역 아파트 가격에 영향을 주는 요인들을 분석하였다. 본 연구의 실증 결과에 의하면, 공급측면의 변수보다는 수요측면의 변수가 서울지역 아파트 가격에 더 크게 영향을 주는 것으로 나타났다. 이러한 결과는 향후 주택가격의 거품 발생과 이에 따른 부작용을 막기 위하여 정부는 보다 선제적인 정책을 취할 필요가 있다는 것을 시사하고 있다. 예를 들어, 단기의 투기적 수요를 억제할 수 있는 세제 정책과 투기적 환경이 조성되지 않도록 하는 이자율 정책 등이 필요할 것으로 판단된다.

주요용어 : 서울 아파트 가격, 거시경제변수, 매수우위지수, 주택담보대출 이자율.

### 1. 서론

본 연구는 주택가격에 영향을 줄 수 있는 거시경제 변수들을 고려하여, 2008년 세계금융위기 이후부터 최근까지 서울지역 아파트가격 변화의 원인에 대하여 분석하고자 한다. 단기간에 주택 가격에 거품이 발생하는 가장 큰 원인은 공급측면보다는 수요측면에 그 원인이 있을 가능성이 매우 크다. 이러한 점을 고려하여 본 연구는 수요측면의 변수로서 이자율을 포함하였다. 단기 투기적 수요변수를 고려하여 주택매수에 대한 기대심리 변수를 수요측 변수에 추가하였다. 또한 주택건축 허가면적을 공급측 변수로 그리고 경제성장률을 통제변수로 추가하였다.

이러한 변수들 사이에 장기적으로 공적분 관계가 성립하는지 살펴본 후, 최종적으로 이러한 변수들이 어떻게 주택가격에 영향을 주는지 분석하였다. 실질적으로 수요측 변수들이 주택가격에 영향을 주었다면, 이자율 및 주택매수 기대심리 등과 같은 수요측 변수들이 주택가격 거품의 원인이라는 것을 암시한다. 이에 따라, 본 연구의 결과는 향후 주택시장의 안정화를 위한 정책적 시사점을 제공한다.

### 2. 자료 및 결과 분석

본 연구는 2007년 4분기부터 2022년 4분기 자료를 이용하였다. GDP와 같이 월별 자료가 제공되지 않는 시계열이 있기 때문에 분기별 자료에 기초하였으며, 월별 자료는 평균값을 사용하여 분기별 자료로 변환하였다. 이자율을 제외한 모든 자료는 로그값을 사용하였다. 수요측면의 자료로 주택담보대출금리 그리고 단기 투기적 수요를 고려하기 위하여 서울지역 아파트 매수우위지수를 포함하였다. 이 두 자료는 국민은행 통계자료를 이용하였다. 공급측면의 자료로 한국부동산원이 제공하는 서울지역의 민간아파트 분양허가실적을 이용하였다. 거시경제 기초여건을 반영하는 통제변수

<sup>1</sup>48513 부산광역시 남구 용소로 45, 부경대학교 경제학과 교수. E-mail: shsim@pknu.ac.kr

로 한국은행에서 제공하는 경제성장률(GDP 성장률) 자료를 이용하였다.

주택가격으로는 주택시장에서 가장 큰 비중을 차지하고 있는 아파트 가격을 채택하였다. 다른 지역보다 서울지역의 아파트 가격의 변동폭이 가장 컸으며, 이는 서울 아파트 가격에 거품이 가장 많이 발생했다는 것을 암시하고 있다. 실제로 기존 연구들(Sim, 2020, 2021)에서도 서울지역 아파트 가격이 다른 지역에 비해 거품이 가장 많이 낀 것으로 보고하고 있다. 따라서 본 연구는 서울지역 아파트 가격의 변동원인에 대하여 분석하고자 한다. 일반적으로 국민은행에서 제공하는 호가 위주의 주택가격 지수는 실거래가격 지수보다 평활화(smoothing) 문제가 발생하는 것으로 알려져 있다(Sim, 2015). 따라서 주택가격의 움직임을 보다 정확하게 반영하기 위하여 본 연구는 한국부동산원이 제공하는 서울 아파트 실거래가격지수를 이용하였다.

OLS 기법을 이용하여 모수를 추정하기 위해서는 개별 자료의 시계열 성격을 먼저 규명해야 한다. 즉, 허구적 회귀(spurious regression)의 원인이 되는 각 변수의 불안정성에 대한 검정이 먼저 필요하다. 본 연구에서는 불안정성을 확인하기 위한 단위근 검정으로 ADF(Augmented Dickey-Fuller) 및 PP(Philips-Perron, 1988) 검정을 이용한다. Table 1의 단위근 검정결과에 의하면, 수준변수의 경우 GDP 성장률을 제외한 모든 변수들에 단위근이 존재하는 것으로 나타났다. 따라서 대부분의 수준 변수들은 불안정한 (non-stationary) 시계열인 것을 알 수 있다. 한편 수준변수들을 1차 차분한 결과, 대부분의 경우 1% 유의수준에서 단위근이 존재한다는 귀무가설이 기각되기 때문에 1차 차분한 변수들은 안정적인 시계열이라고 할 수 있다.

Table 1. Unit root test

	HPI	PERMIT	Interest Rate	BUYINDEX	GDP
Level					
ADP	-0.858	-2.212	-2.339	-1.731	-7.048***
PP	-0.113	-2.325	-2.055	-1.717	-7.024***
First Difference					
ADP	-3.640***	-6.773***	-6.007***	-7.078***	-8.047***
PP	-3.502**	-7.016***	-7.016***	-7.053***	-34.31***

Note: \*\*\* and \*\* indicate 1% and 5% significance levels, respectively.

Table 2. Cointegration test

rank	$\lambda_{trace}$	$\lambda_{max}$
$r = 0$	0.4525**	0.4525**
$r \leq 1$	0.3876**	0.3876**
$r \leq 2$	0.2309	0.2309
$r \leq 3$	0.1196	0.1196
$r \leq 4$	0.0490	0.0490

Note: \*\* indicates 5% significance level.

Table 3. Estimation result

PERMIT	-0.1230
Interest Rate	-0.0963***
BUYINDEX	0.1429***
GDP	-0.0398
$R^2$	0.35
Durbin-Watson	1.27

Note: \*\* indicates 1% significance level.

모든 시계열 자료들이 I(1) 과정을 따르는 것으로 나타났기 때문에 변수들 간의 장기균형관계를 파악하기 위하여 공적분(cointegration) 분석을 시행하였다. 본 연구에서는 Johansen(1988) 공적분 검정식에 상수항을 포함하였다. Table 2의  $\lambda_{trace}$  및  $\lambda_{max}$  값에 의거하여 공적분 식이 두 개 존재하는 것으로 나타났다. 따라서 변수들이 불안정한 시계열일지라도 변수들 간에 공적분 관계가 형성되므로 수준변수를 이용하여 OLS 모형을 추정할 수 있다.

Table 3은 장기적으로 각 변수들이 어떻게 주택가격에 영향을 주는지 보여주고 있다. 먼저 아파트 분양허가실적(PERMIT)은 예상대로 주택가격과 음의 관계를 보여주고 있으나 통계적으로 의미가 없는 것으로 나타났다. 이 결과에 따라 본 연구의 샘플 기간 동안 주택가격은 공급측면에 크게 영향을 받지 않는 것으로 해석할 수 있다.

경제성장(GDP)도 마찬가지로 주택가격에 크게 영향을 주지 않는 것으로 나타났다. 한편 수요측면의 변수인 아파트 매수우위지수(BUYINDEX)의 경우 주택가격에 강하게 영향을 주는 것으로 나타났다. 즉, 매수 의사가 높을수록 아파트 가격이 상승한다고 볼 수 있으며, 이는 단기적인 투기 수요가 주택가격을 크게 상승시킨다는 것을 의미한다. 주택담보대출 금리(Interest Rate) 역시 주택가격에 크게 영향을 주는 것으로 나타났다. 이자율 추정값은 예상대로 음의 값을 보여주는데, 이는 본 연구의 샘플기간 동안 이자율과 주택가격 간의 지속적인 음의 관계를 뒷받침하는 결과라고 볼 수 있다.

### 3. 결론

최근 몇 년 동안 서울의 주택가격은 크게 상승하였으며, 이는 주택가격에 거품이 발생한 것으로 기존연구들에 의해 밝혀지고 있다. 이에 본 연구는 2008년 세계금융위기 이후 서울의 아파트 가격에 영향을 주는 요인들을 분석하였다.

본 연구의 결과에 따르면, 공급측면보다는 수요측면의 변수들이 주택가격에 크게 영향을 준 것으로 나타났다. 최근 몇 년 동안 갑작스런 인구증가나 소득증가와 같이 실거주 위주의 주택수요를 증가시키는 요인들은 발견되지 않는다. 따라서 주택 공급이 부족하기 때문에 주택 가격이 갑자기 급등하였다는 주장은 논리적으로 설득력이 떨어진다. 이와는 반대로, 본 연구의 실증결과처럼, 단기의 투기적 수요나 이러한 투기적 수요를 뒷받침해 줄 수 있는 경제 환경이 주택가격을 급등시킨 원인이라는 것을 알 수 있다.

장기적으로 기록적인 낮은 이자율은 시장에 유동성을 지속적으로 증가시켰으며, 이는 부동산 시장으로 투자자금을 유입시키게 된 가장 큰 요인으로 볼 수 있다. 이러한 상황에서, 주택시장에서 큰 투자수익을 목격한 많은 사람들이 주택에 대한 투자를 늘림으로써 주택가격이 크게 상승한 것으로 볼 수 있다. 다시 말하면 주택에 대한 투기적 수요가 결정적으로 주택가격을 급등시킨 요인이라고 할 수 있다. 본 연구의 결과에서도 보듯이, 우리나라 주택시장 중 특히 서울의 경우 단기 투기적 수요에 의해 주택가격이 갑자기 상승하기 쉬운 여건에 놓여있음을 알 수 있다. 이를 방지할 경우, 외적인 충격이 발생하지 않는 한 주택가격에 거품이 낄 수 있으며, 이에 따라 거품이 지속되거나 거품이 터질 경우 국내 경제에 미치는 부정적인 영향은 막대할 것이다.

단기적으로는 공급측면의 요인보다는 수요측면의 요인이 주택가격에 더 크게 영향을 주기 때문에, 단기의 투기적 수요를 억제할 수 있는 세계 정책과 투기적 환경이 조성되지 않도록 하는 이자율 정책 등이 필요할 것으로 판단된다. 이러한 정책들은 경제에 부정적인 영향을 미칠 수도 있지만, 거품과 관련된 부정적인 파급효과를 막을 수 있는 긍정적인 정책 효과가 더 클 경우 수요억제 정책은 매우 중요한 정책수단이 될 것으로 판단된다.

## References

- Johansen, S. (1988). Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.
- Sim, S. H. (2015). The relationship between transaction-based prices and appraisal-based prices in Korean Housing Markets, *Journal of the Korean Data Analysis Society*, 17, 2579-2593 (in Korean).
- Sim, S. H. (2020). Bubbles and spillovers in Korean housing markets, *Journal of the Korean Data Analysis Society*, 22, 483-500.
- Sim, S. H. (2021). Testing for house price bubble in Seoul, Korea, *Journal of the Korean Data Analysis Society*, 23(5), 2025-2037.

## The main factors influencing the house prices in Seoul

*Sunghoon Sim*

### Abstract

This study analyzes the main factors that influence the apartment prices in Seoul during 2007:Q4~2022Q4. Our empirical findings show that the demand side factors have more impact on the house prices than the supply side factor. This result implies that the policy makers need to take preemptive measures in order to curb negative effects caused by house price bubbles. The policies may include both the tax policy to prevent short-run speculative demand and the interest rate policy to avoid the speculative circumstances caused by huge amount of liquidity into the housing markets.

*Keywords* : APT price, macroeconomic factors, buying-index, housing loan interest rate.

## 토픽모델링을 활용한 기술동향 파악: AI 관련특허를 중심으로

김용우<sup>1</sup>, 김민구<sup>2</sup>, 김정희<sup>3</sup>

### 요약

인공지능(AI) 기술은 최근 빠른 속도로 발전하면서 다양한 분야에 변화를 가져오고 있다. AI가 기존 산업을 혁신하거나 새로운 산업을 창출하는 구실을 하게 되면서 AI 관련 기술의 동향 파악은 더욱 중요해지고 있다. 본 연구에서는 주요국의 AI 관련 특허를 정의하고 수집한 후 토픽모델링을 진행하여 AI 관련 특허를 몇 개의 주제로 나누었다. 또한, 나누어진 주제분류에 기반하여 주요국(미국, 유럽연합, 중국, 한국)의 특허출원 현황을 살펴보았다. 그 결과, 연구개발에 집중하고 있는 분야의 차이가 있음을 확인하였으며, 이러한 상황에서의 우리나라 AI 연구개발 방향에 대해서도 논의하였다.

주요용어 : 특허, 인공지능, 토픽모델링, LDA.

### 1. 서론

기술예측(Technology forecasting)은 기술혁신 및 산업발전을 위해 중요한 전략적 의사결정을 지원하는 핵심적인 분야 중 하나이다. 기술예측은 기술의 발전 경향성, 시장의 수요 변화, 경쟁환경 등을 종합적으로 분석하여 현재에 어떤 기술이 개발되고 있는지, 그리고 미래에는 어떤 기술이 유망할 것인지 살펴보고, 이를 토대로 기업이나 정부 등이 효율적인 전략을 수립할 수 있도록 도울 수 있다. 과거에는 델파이 기법처럼 경험과 직관에 의존하는 주관적인 예측이 일반적이었다. 그러나 절차의 복잡성, 시간 및 경제적 비용 문제, 소수 전문가의 과도한 의견 발휘에 의한 왜곡 등 많은 단점이 존재한다. 그러나 최근에는 데이터 분석, 인공지능 등의 기술을 활용하여 이전보다 객관적이고 신뢰성이 높은 예측이 가능해졌다. 인공지능(AI) 기술은 최근 빠른 속도로 발전해왔으며, 다양한 분야에서 혁신적인 변화를 가져오고 있다. 이러한 변화는 기존 산업을 혁신하거나 새로운 산업을 창출하는 중요한 역할을 하고 있으며, AI 관련 기술의 동향 파악은 더욱 중요해지고 있다. 특히, AI 기술 분야에서는 최근 많은 특허가 출원되어 있으며, 이를 이용하여 AI 기술 동향을 파악하는 것은 매우 유용하다. 이러한 배경 아래 본 연구에서는 주요국의 AI 관련 특허를 수집 및 분석하고 패턴에 따라서 유형별로 구분한다. 또한, AI 특허의 분야별 트렌드 및 각국의 기술발전 동향에 대해서도 살핀다.

<sup>1</sup>(교신저자) 04763 서울시 성동구 왕십리로 222, 한양대학교 기술경영전문대학원 기술경영학과 박사수료.  
E-mail: ywkim@hanyang.ac.kr

<sup>2</sup>04763 서울시 성동구 왕십리로 222, 한양대학교 기술경영전문대학원 기술경영학과 박사수료.  
E-mail: loginmin9@hanyang.ac.kr

<sup>3</sup>04763 서울시 성동구 왕십리로 222, 한양대학교 기술경영전문대학원 기술경영학과 박사수료.  
E-mail: chocojh@hanyang.ac.kr



## 2. 분석 방법

먼저 신뢰할 수 있는 특허 정보를 추출하고 대상 특허를 정의하는 것이 중요하다. 본 연구에서는 유럽특허청(EPO)에서 제공하는 특허 데이터베이스인 PATSTAT을 사용하여 신뢰성을 높이는 한편, WIPO(세계지식재산권기구)에서 AI 관련 특허를 정의하기 위해 사용한 방법(WIPO, 2019)을 이용하여 AI 관련 특허를 정의하여 관련 특허를 자의적으로 정의할 경우 발생하는 문제를 예방하고자 하였다. 단순 키워드로 특허를 추출하거나, 관련된 특허분류코드에 속한 특허를 사용할 경우 대상 특허가 너무 작게 정의되거나 관련이 없는 특허도 분석대상에 포함되기 때문이다. 2010년부터 2019년까지 출원된 AI 관련특허는 638,174개이며 그중에서 초록의 언어가 영어인 특허인 408,429개의 특허가 본 연구의 분석대상이다. 선정된 AI 관련 특허의 제목과 초록을 수집하여 DTM(Document Term Matrix)형태로 임베딩(Embedding)하고, 특허를 분야별로 나누는 과정에서 LDA(Latent Dirichlet Allocation) 방법을 이용하였다. 전체 프로세스는 Figure 1에 나타나 있다.

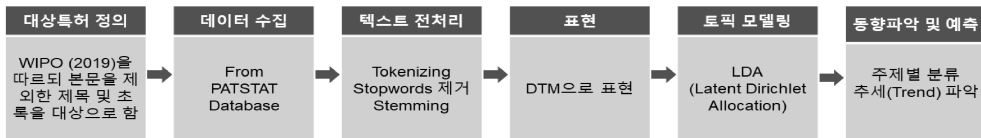


Figure 1. Process

## 3. 분석 결과

대상 특허를 최종적으로 11개의 주제로 분류하였다. 주제 수를 달리하여 LDA 모델을 학습한 후, 군집화 결과를 평가하였다. LDA 모델의 경우 주제 수를 사용자가 직접 사전에 설정해야 한다. 주제 수에 따라 군집화 결과가 달라지며 군집화 결과를 평가할 수 있는 척도로 일반적으로 PPL(Perplexity)이 사용된다. 그러나 학습된 주제의 품질과 PPL간 상관관계가 존재하지 않는 경우도 종종 발생한다(Chang et al., 2009). 본 연구에서는 주제 수 변화에 따른 PPL 값이 유사하여 PPL로 군집의 개수를 판단하기가 어려웠다. 이러한 이유로 결과를 수동으로 직접 평가하였다. 결과 검토 후 11개로 나누어진 주제의 이름을 연구자가 직접 생성하였으며, 토픽별 특허의 분포는 Table 1과 같다. 최근의 국가별 특허출원 추이는 Figure 2와 같으며, 특허 최근의 중국의 특허출원이 두드러짐을 확인할 수 있다. 최근 AI 관련 특허의 홍수는 대부분 중국으로부터 야기된 것으로 볼 수 있다.

주제별로는 [2. User Interface] 및 [8. Data Processing System] 관련 특허의 출원이 주춤한 모양새이다. 반대로 [7. Calculation and Optimization Process], [6. Image / Face Recognition], [5. Connecting Module & Equipment] 관련 특허의 경우 출원이 증가하고 있다. 나머지 주제들의 출원 추세에 대해서도 Figure 3에서 확인할 수 있다.

3년간(2016-2019)의 주요국 출원 특허 비중을 주제별로 나타낸 그림이 Figure 4에 나타나 있다. 미국 및 유럽의 경우 특허출원 패턴이 매우 유사함을 알 수 있다. 출원된 특허 대부분은 [8. Data Processing System] 및 [2. User Interface] 관련 영역에 집중되어 있다. 중국의 경우 출원패턴이 미국이나 유럽과는 다른데 주로 [5. Connecting Module & Equipment], [6. Image & Face Recognition], 그리고 [7. Calculation & Optimization Process]에 집중되어 있다. 한국의 경우 중국보다는 미국 및 유럽과 유사한 형태의 출원 행태를 보이고 있다.

Table 1. Patent distribution by topic

Topic Number	Topic	Number of Patents	Proportion (%)
2	User Interface (UI)	102,363	25.1%
8	Data Processing System (Intelligent System)	69,441	17.0%
6	Image / Face Recognition	63,976	15.7%
7	Calculation and optimization Process	54,547	13.4%
5	Connecting Module & Equipment	40,356	9.9%
10	Driving Control	23,818	5.8%
1	Sensor Signal and Information processing (SSIP)	21,740	5.3%
0	Machine Learning (ML)	12,122	3.0%
3	Positioning & Locating	8,969	2.2%
9	Networking	6,978	1.7%
4	Audio & Sound	4,119	1.0%

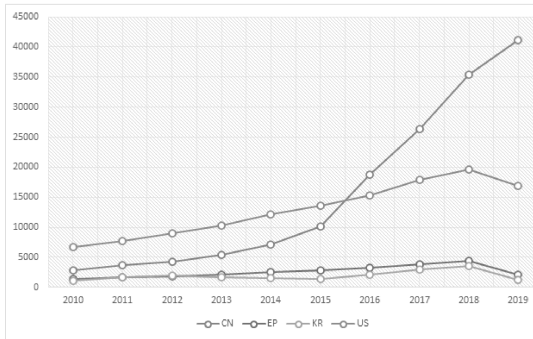


Figure 2. Patent Application Trends by Country

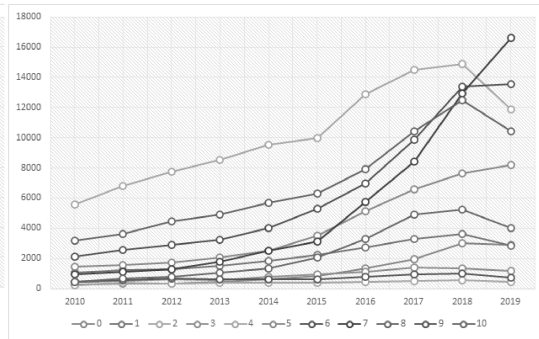


Figure 3. Patent Application Trend by Topic

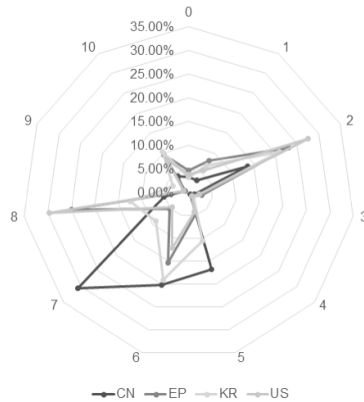


Figure 4. Patent Application by Topic (2016-2019)

중국은 인공지능 산업을 국가적으로 중요전략산업으로 지정하고 이에 따라 인공지능 관련 기업에 대한 지원은 물론 인재 양성을 위해서도 힘쓰고 있다. 또한, 중국의 많은 인구는 대규모 데이터 수집 및 활용에 있어서 유리하다. 이러한 정책적·환경적 이점 덕분에 최근에 주목받고 있는 이미지 및 얼굴인식 분야에서 많은 연구성과가 있는 것으로 보인다. 중국의 경우 조세우대, 재무금융 등 기술개발 환경 개선·제공에 매우 적극적인 편이며, 세부적으로는 컴퓨터 비전, 자연어처리 영역에 중국 정부의 지원이 집중되어 있다(Kim, 2020). 우리나라 역시 다양한 정책을 통해서 육성하고

있으나, 최근 AI 연구개발의 대형화 추세를 고려했을 때 주요국과의 경쟁이 쉽지 않은 상황이다. 주요국과 다른 영역의 기술개발을 통해 산업에 파급효과를 낼 수 있는지 점검하고 우리만의 기술 영역을 육성하는 것을 고려해볼 수 있다. 동시에 (Kim, 2016)에서 제안한 것처럼 국제공동 연구 협력을 추진을 통해 파급력이 큰 원천기술은 적시에 적용할 수 있도록 준비가 필요하다.

#### 4. 결론

본 연구에서는 최근 주목받는 AI 관련 기술개발 동향을 살펴보기 위해 특허 데이터를 활용하였다. AI 관련 특허를 정의하고 추출한 후 LDA를 사용하여 11가지 주제로 분리하였다. 이후 주제영역별 특허의 출원 추이를 살펴보고, 국가별로 비교하였다. 그 결과 2015년 이후 AI 관련 특허의 출원이 급격하게 증가하는 것을 확인할 수 있었다. 또한, 주요국의 특허출원 양상이 다른 점도 확인하였다. 특히 최근 주목받고 있는 얼굴 및 이미지 인식 분야에서 중국의 연구개발 행보가 두드러졌다. 그러나 모든 특허가 성공적으로 가치를 창출하지는 못한다. 즉, 특허의 양(stock)뿐만 아니라 특허의 질 역시 기술평가의 주요한 요소이므로 이러한 결과로 중국의 AI 기술 수준이 높다고 판단하기는 어렵다. 학술논문 발표량에서도 중국이 세계 1위를 차지했지만, 논문의 질은 미국에 큰 격차로 뒤지는 것으로 나타났기 때문이다(Kim, You, 2022). 시일이 지난 후 특허의 수명이나 특허의 전방인용(Forward Citation) 등 특허의 가치를 측정하는 다른 정보를 통해 각국 AI 기술의 가치를 알아보는 연구가 진행될 필요가 있다.

#### References

- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J., & Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 22.
- Kim, B. W. (2016). Trend Analysis and National Policy for Artificial Intelligence. *Informatization Policy*, 23(1), 74-93. (in Korean).
- Kim, J. J. (2020). China's Artificial Intelligence Technique and Comparison on US-China Artificial Intelligency Policy. *Journal of China Area Studies*, 7(1), 67-97. (in Korean).
- Kim, S, Yoo, H. (2022). The Status of Artificial Intelligence and National Strategy in China during the Xi Jinping era. *Journal of International Relations*, 25(1), 241-273. (in Korean).
- WIPO (2019). WIPO Technology Trends 2019 - Artificial Intelligence. Geneva, Switzerland : World Intellectual Property Organization.

## 2단계 집락 조건부 이산양적모형\*

이기성<sup>1</sup>, 홍기학<sup>2</sup>, 손창균<sup>3</sup>, 박근화<sup>4</sup>, 홍성준<sup>5</sup>

### 요 약

본 논문에서는 집락으로 구성된 모집단에서 얻고자 하는 정보가 이산인 양적 민감한 속성일 때 덜 민감한 속성과 강요질문으로 이루어진 Loynes의 확률장치를 이용하여 “예”라고 응답한 사람들에게만 이산양적모형인 Liu-Chow 모형을 사용하도록 하는 조건부 이산양적모형에 2단계 집락추출법을 적용한 2단계 집락 조건부 이산양적모형을 제안하였다.

주요용어 : 민감한 이산 양적 정보, 덜 민감한 속성, 2단계 집락추출, 조건부 모형.

### 1. 서론

Warner(1965)는 두 개의 민감한 질문과 관련된 질문으로 이루어진 확률장치를 통해 민감한 정보를 획득할 수 있는 확률화응답모형을 제안하였다. Loynes(1976)는 민감한 질문과 강요응답 “예”로 구성된 확률장치를 이용하는 강요질문모형을 제안하여 민감한 질적속성을 추정하였고, Liu, Chow. (1976)는 민감한 이산 양적속성을 추정할 수 있는 이산양적 확률화응답모형을 제안하기도 하였다. 그리고 Carr, Marascuilo.(1982)은 덜 민감한 속성과 강요응답으로 구성된 확률장치로부터 “예”라고 응답한 응답자들만을 대상으로 Loynes의 강요질문모형을 이용하도록 하는 조건부 확률화응답모형을 제안하였다. 본 논문에서는 집락으로 구성된 모집단에서 얻고자 하는 정보가 이산인 양적 민감한 속성일 때 덜 민감한 속성과 강요질문으로 구성된 Loynes의 확률장치를 이용하여 “예”라고 응답한 사람들에게만 Liu-Chow 모형을 사용하도록 하는 조건부 이산양적모형에 2단계 집락추출법을 적용한 2단계 집락 조건부 이산양적모형을 제안하고자 한다.

### 2. 2단계 집락 조건부 이산양적모형

우리가 관심을 가지고 있는 모집단은 각 집락 크기가  $M_i (i = 1, 2, \dots, N)$  인  $N$ 개의 집락으로 구성되어 있다고 가정하자.  $N$ 개의 집락 중에서  $n$ 개의 집락을 단순임의비복원으로 추출한 후에 추

\*이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (2022R1F1A1063263).

<sup>1</sup>55338 전북 완주군 삼례읍 삼례로 443 우석대학교 아동사회복지학부 교수. E-mail: gisung@woosuk.ac.kr

<sup>2</sup>58245 전남 나주시 건재로 185 동신대학교 컴퓨터학과 교수. E-mail: khhong@dsu.ac.kr

<sup>3</sup>38066 경북 경주시 동대로 123 동국대학교 빅데이터·응용통계학과 교수. E-mail: ckson85@dongguk.ac.kr

<sup>4</sup>07511 서울특별시 강서구 금남화로 154 한국문화관광연구원 선임전문위원. E-mail: pkhkwen@kcti.re.kr

<sup>5</sup>13822 경기도 과천시 경마공원대로 107 마사회 말산업연구소 연구원. E-mail: hsj8129@naver.com

출된 각 집락으로부터 다시  $m_i$  ( $i = 1, 2, \dots, n$ ) 개의 조사단위를 단순임의복원추출하는 2단계 집락추출법을 조건부 이산양적모형에 적용해 보고자 한다. 첫 번째 단계에서  $m_i$  명의 응답자에게 아래와 같은 강요질문모형의 확률장치를 이용하여 선택된 질문에 대하여 응답하게 한다.

	질문내용	선택확률
질문 1	당신께서는 덜 민감한 속성 $B_i$ 가 있으십니까?	$p_i$
질문 2	‘예’라고 응답하십시오.	$1 - p_i$

이때, 응답자들은 확률장치를 통해 질문 1이 나타나면 “예” 또는 “아니오”라고 응답하고, 질문 2가 나타나면 “예”라고 응답하게 된다.  $i$  번째 집락의 첫 번째 단계에서 응답자가 완전히 진실된 응답을 하는 경우 “예”라고 응답할 확률  $l_{i1}$ 는 아래와 같다.

$$l_{i1} = p_i \theta_{i1} + (1 - p_i). \quad (2.1)$$

여기서,  $\theta_{i1}$ 은  $i$  번째 집락의 덜 민감한 속성  $B_i$ 의 모비율이다.

그리고 “예”라고 응답한 사람들의 수를  $m_i'$  명이라 하면,  $\hat{l}_{i1} = \frac{m_i'}{m_i}$  이 되므로 덜 민감한 속성  $B_i$ 에 대한 모비율  $\theta_{i1}$ 의 추정량  $\hat{\theta}_{i1}$ 는 아래와 같다.

$$\hat{\theta}_{i1} = \frac{\hat{l}_{i1} - (1 - p_i)}{p_i} = \frac{1}{m_i p_i} [m_i' - (1 - p_i) m_i]. \quad (2.2)$$

$i$  번째 집락의 두 번째 단계에서는 첫 번째 단계에서 “예”라고 응답한 사람들  $m_i'$  명만이 Liu-Chow 모형의 확률장치를 통해 선택된 질문에 대해 응답하도록 한다. Liu-Chow 모형의 확률장치는 붉은 색 공과 파란 색 공으로 구성되어 있다. 붉은 색 공에는 여러 개의 민감한 질문 내용이 쓰여져 있고, 파란 색 공에는 0부터  $s_i$ 까지의 정수가 미리 정해진 비율에 따라 쓰여져 있다. 응답자들은 확률장치에 있는 공들을 잘 섞은 뒤에 선택된 공에 쓰여져 있는 내용을 응답하면 된다. 만약 붉은 색 공이 나오면 응답자가 가지고 있는 이산인 민감한 양적속성의 값을 응답하며, 파란 색 공이 나오면 공에 쓰여진 숫자를 응답한다. 이때, 파란 색 공에 쓰여진 숫자들은 민감한 이산인 양적속성의 값들을 포함하고 있다.  $i$  번째 집락에서  $t_i$  ( $t_i = 0, 1, 2, \dots, s_i$ ) 라는 숫자가 적혀진 파란 색 공의 수를  $b_{it}$ 라 하고, 민감한 질문이 적혀져 있는 붉은 색 공의 수를  $r_i$ 라 하면, 전체 공의 수는  $r_i + b_i$ 가 되며, 이 때  $b_i = \sum_{t_i=0}^{s_i} b_{it}$ 이다. 따라서  $i$  번째 집락의 응답자들이 첫 번째 단계에서 “예”라고 응답했다는 조건하에 두 번째 단계에서  $t_i$  ( $t_i = 0, 1, 2, \dots, s_i$ ) 라고 응답할 확률  $\lambda_{it}$ 는 아래와 같다.

$$\lambda_{it} = \frac{\pi_{it}}{l_{i1}} \left( \frac{r_i}{r_i + b_i} \right) + \frac{b_{it}}{r_i + b_i}. \quad (2.3)$$

여기서,  $\pi_{it}$ 는  $i$  번째 집락의 민감한 이산양적속성  $t$ 에 대한 모비율이다.

$i$  번째 집락의 응답자들  $m_i'$  명 중에서  $t_i$  라고 응답한 응답자의 수를  $m_{it}$ 라 두면  $\hat{\lambda}_{it} = \frac{m_{it}}{m_i'}$ 가 되므로  $t_i$ 에 대한 모비율  $\pi_{it}$ 의 추정량  $\hat{\pi}_{it}$ 는 아래와 같다.

$$\hat{\pi}_{it} = \hat{l}_{i1} \left( \hat{\lambda}_{it} - \frac{b_{it}}{r_i + b_i} \right) / \left( \frac{r_i}{r_i + b_i} \right) = \frac{1}{m_i \left( \frac{r_i}{r_i + b_i} \right)} \left[ m_{it} - \left( \frac{b_{it}}{r_i + b_i} \right) m_i' \right]. \quad (2.4)$$

이때,  $\hat{\pi}_{it}$ 의 기댓값은  $E(\hat{\pi}_{it}) = \pi_{it}$ 이고, 그 분산은 다음과 같다.

$$V(\hat{\pi}_{it}) = \frac{l_{i1}}{m_i \left( \frac{r_i}{r_i + b_i} \right)^2} \left[ \lambda_{it} (1 - l_{i1} \lambda_{it}) + \left( \frac{b_{it}}{r_i + b_i} \right)^2 (1 - l_{i1}) - 2 \left( \frac{b_{it}}{r_i + b_i} \right) (1 - l_{i1}) \lambda_{it} \right]. \quad (2.5)$$

한편, 민감한 이산양적속성  $t$ 에 대한 모비율  $\pi_t$ 는 다음과 같이 표현할 수 있다.

$$\pi_t = \frac{1}{M_0} \sum_{i=1}^N M_i \pi_{it}, \quad M_0 = \sum_{i=1}^N M_i \quad (2.6)$$

$i$  ( $i = 1, 2, \dots, n$ ) 번째 집락으로부터 응답자들이 단순임의복원으로 추출되었을 때,  $t$ 에 대한 모비율  $\pi_t$ 의 추정량  $\hat{\pi}_t$ 는 다음과 같이 정의할 수 있다.

$$\hat{\pi}_t = \frac{N}{nM_0} \sum_{i=1}^n M_i \hat{\pi}_{it} = \frac{1}{nM} \sum_{i=1}^n M_i \hat{\pi}_{it}, \quad \bar{M} = \frac{M_0}{N} \quad (2.7)$$

<정리 1>  $\hat{\pi}_t$ 는  $\pi_t$ 의 비편향추정량이 된다.

<정리 2>  $N$ 개의 집락에서  $n$ 개의 집락을 단순임의비복원으로 추출하고, 추출된 집락으로부터 다시  $m_i$ 개의 조사단위를 단순임의복원으로 추출할 경우  $\hat{\pi}_t$ 의 분산은 다음과 같다.

$$V(\hat{\pi}_t) = \frac{N-n}{nN(N-1)} \sum_{i=1}^N \left( \frac{M_i \pi_{it}}{\bar{M}} - \pi_t \right)^2 + \frac{1}{nNM^2} \sum_{i=1}^N M_i^2 \frac{l_{i1}}{m_i \left( \frac{r_i}{r_i + b_i} \right)^2} \left\{ \lambda_{it} (1 - l_{i1} \lambda_{it}) + \left( \frac{b_{it}}{r_i + b_i} \right)^2 (1 - l_{i1}) - 2 \left( \frac{b_{it}}{r_i + b_i} \right) (1 - l_{i1}) \lambda_{it} \right\}. \quad (2.8)$$

### 3. 효율성 비교

이 절에서는 2단계 집락 조건부 이산양적모형과 2단계 집락추출법에 의한 Liu-Chow 모형과의 효율성을 비교해 보고자 한다. 우선, Liu-Chow 모형에 2단계 집락추출법을 적용하여 민감한 이산양적속성  $t$ 에 대한 모비율  $\pi_t$ 의 추정량  $\hat{\pi}_t$ 의 분산을 도출해 보면 다음과 같다.

$$V_{LC}(\hat{\pi}_t) = \frac{N-n}{nN(N-1)} \sum_{i=1}^N \left( \frac{M_i \pi_{it}}{\bar{M}} - \pi_t \right)^2 + \frac{1}{nNM^2} \sum_{i=1}^N M_i^2 \frac{1}{m_i \left( \frac{r_i}{r_i + b_i} \right)^2} L_{it} (1 - L_{it}). \quad (3.1)$$

여기서,  $L_{it} = \pi_{it} \left( \frac{r_i}{r_i + b_i} \right) + \frac{b_{it}}{r_i + b_i}$  이다.

$N = 4$  개의 집락에서  $n = 2$  개의 집락을 추출한 다음, 추출된 각 집락에서 다시  $m_i$  개의 조사단위를 추출한다고 가정하자. 이 때,  $M_i$  의 크기는  $M_1 = 1,000, M_2 = 2,000, M_3 = 3,000, M_4 = 4,000$  이고  $m_i$  의 크기는  $m_1 = 100, m_2 = 200, m_3 = 300, m_4 = 400$  라고 하자.  $\pi_{it}$  는  $\pi_{10} = 0.1, \pi_{11} = 0.2, \pi_{12} = 0.3, \pi_{13} = 0.4, \pi_{20} = 0.15, \pi_{21} = 0.25, \pi_{22} = 0.3, \pi_{23} = 0.3, \pi_{30} = 0.2, \pi_{31} = 0.3, \pi_{32} = 0.4, \pi_{41} = 0.25, \pi_{42} = 0.4, \pi_{43} = 0.3$  이라 가정하자. 그리고  $i$  번째 집락의 확률장치에는 민감한 질문이 쓰여진 있는 붉은 색 공  $r_i$  개와  $t_i$  ( $t_i = 0, 1, 2, 3$ ) 라는 숫자가 쓰여진 파란 색 공  $b_{it}$  개가 있다고 가정하자. 이 때,  $r_i = 4, b_{i0} = b_{i1} = b_{i2} = b_{i3} = 1$  이라고 가정하자. 그리고 2단계 집락 조건부 이산양적모형의 첫 번째 단계에서 덜 민감한 속성  $B_i$  의 모비율  $\theta_{i1}$  값을 변화시켜가면서 2단계 집락 조건부 이산양적모형의 분산 식(2.8)과 2단계 집락 Liu-Chow 모형의 분산 식(3.1)을 사용하여 상대효율 ( $RE = \frac{V_{LC}(\hat{\pi}_t)}{V(\hat{\pi}_t)}$ )을 다음 Table 1과 같이 구할 수 있다. 이 때 상대효율이 1보다 크게 나타난 경우 2단계 집락 조건부 이산양적모형이 2단계 집락 Liu-Chow 모형보다 효율적임을 의미한다.

$p_i$  값이 커질수록 2단계 절차를 거친 2단계 집락 조건부 이산양적모형이 2단계 집락 Liu-Chow 모형보다 효율성이 높게 나타남을 알 수 있었다. 또한  $\theta_{i1}$  값이 1에 가까워질수록 2단계 집락 조건부 이산양적모형과 2단계 집락 Liu-Chow 모형의 효율성이 비슷해짐을 알 수 있었다.

Table 1. Efficiency comparison

$p_i$	$\theta_{i1}$	$t_i$			
		0	1	2	3
0.1	0.5	1.0374	1.0300	1.0268	1.0295
	0.6	1.0297	1.0239	1.0213	1.0234
	0.7	1.0221	1.0178	1.0159	1.0175
	0.8	1.0146	1.0118	1.0105	1.0116
	0.9	1.0072	1.0058	1.0052	1.0057
0.3	0.5	1.1213	1.0960	1.0850	1.0941
	0.6	1.0948	1.0754	1.0668	1.0739
	0.7	1.0694	1.0555	1.0493	1.0544
	0.8	1.0452	1.0363	1.0323	1.0356
	0.9	1.0221	1.0178	1.0159	1.0175

### Reference

Carr, J. W., Marascuilo, L. A. (1982). Optimal randomized response models and methods for hypothesis testing, *Journal of Educational Statistics*, 7, 295-310.

Liu, P. T., Chow, L. P. (1976). A new discrete quantitative randomized model, *Journal of the American Statistical Association*, 71, 72-73.

Loynes, R. M. (1976). Asymptotically optimal randomized response procedures, *Journal of the American Statistical Association*, 71, 924-928.

Warner, S. L. (1965). Randomized response ; A survey technique for eliminating evasive answer Bias, *Journal of the American Statistical Association*, 60, 63-69.

## 자료 포락 분석(DEA)을 이용한 기술 고효율 스타트업 분류모델: 민간주도형 기술창업지원 팀스(TIPS) 선정 기업을 중심으로

김정호<sup>1</sup>, 박현민<sup>2</sup>, 오주희<sup>3</sup>

### 요 약

본 연구는 정보의 부족함과 불확실성 속에서 초기 스타트업의 성과를 객관적으로 측정하고 평가하는 새로운 지표를 개발하였다. 초기 스타트업의 효율성 지표를 도출하기 위해 크런치베이스(Crunchbase)와 KIND 데이터베이스에서, 스타트업 활동과 관련된 기술 사용변수, 특히, 크런치베이스 랭킹 등 다양한 투입/산출 변수에 관한 정보를 수집하여, 효율성이 높은 초기 스타트업을 식별하는 방법을 제안하였다. 이를 위해, 회귀분석과 랜덤포레스트(RandomForest) 및 부스팅(XGBoost) 기반 다양한 머신러닝 기법을 활용하여 효율성기준 상위 10, 30, 50 백분위수에 속하는 스타트업을 분류하는 예측 모델을 개발하였다. 초기 스타트업의 효율성지표 분류 모델을 통해, 기술효율성 지표기준 상위 백분위수로 올라갈수록 예측 정확도가 높은 것을 확인할 수 있었다. 이 연구의 결과는, 특히 정보가 부족하고 불확실한 초기 시장에서 스타트업의 성과를 평가하고 예측하는데 큰 도움이 될 것이다. 향후 이 연구에서 개발된 분류 모델과 지표를 활용하여, 초기 스타트업의 성장 가능성과 지속성을 평가하는 데 초점을 맞춰야 할 것이다. 이러한 접근 방식은 투자자들이 데이터를 기반으로 더욱 정확한 투자 결정을 내릴 수 있도록 도와줄 것이다. 이 연구의 결과는 초기 스타트업 생태계의 이해와 발전에 중요한 도움을 줄 것으로 기대된다.

주요용어 : 스타트업, 기술효율성, 머신러닝, 팀스(TIPS), 자료포락분석, 크런치베이스.

---

<sup>1</sup>37554 경상북도 포항시 북구 흥해읍 한동로 558, 한동대학교 ICT창업학부 학생.  
E-mail: jeonghokim123@gmail.com

<sup>2</sup>37554 경상북도 포항시 북구 흥해읍 한동로 558, 한동대학교 ICT창업학부 학생.  
E-mail: gusalsgmlldud@handong.ac.kr

<sup>3</sup>(교신저자)37554 경상북도 포항시 북구 흥해읍 한동로 558, 한동대학교 경영경제학부 교수,  
E-mail: jooheeeoh@handong.edu



## An Acoustic Feature Based Popularity Prediction Using Machine Learning Approach: Indie Music Scene

*Eunbi Cho*<sup>1</sup>, *Haeun Jang*<sup>2</sup>, *Joohee Oh*<sup>3</sup>

### Abstract

In the rapidly evolving landscape of the music industry, predicting the popularity of songs has become a crucial task for the songs producers and record industry executors. Especially the unique characteristics of the indie music scene call for a distinct approach. This research paper presents a novel machine learning-based methodology for popularity prediction in the indie music scene, focusing specifically on acoustic features of songs extracted from 2017 to 2023 monthly top 100 Melon chart. In this study, we employed various machine learning techniques, such as Logistic Regression, Extreme Gradient Boosting (XGBoost), Gradient Boosting, and Random Forest, to analyze and predict the success of album sales. Among these techniques, the XGBoost algorithm demonstrated the highest performance, achieving an F1 Score of 0.8 in the integrated sampling of acoustic features and interactive features. To gain insights into the factors driving album success, we utilized the SHAP (SHapley Additive exPlanations) technique. Through the application of SHAP, we discovered timbral features play a crucial role in significantly influencing the prediction of a song's success. This comprehensive approach allowed us to gain a deeper understanding of the intricate dynamics underlying album success and empowering us to provide better-informed decisions for indie musicians to enhance the chances of success in music industry.

*Keywords* : Music Popularity Prediction, Indie Music, Acoustic Features, Machine Learning.

---

<sup>1</sup>Undergraduate Student, School of Management and Economics, Handong Global University, Handongro 558, 37554, Korea, Email: 22000724@handong.ac.kr

<sup>2</sup>Undergraduate Student, School of Communication Arts, Handong Global University, Handongro 558, 37554, Korea, Email: jjangha@handong.ac.kr

<sup>3</sup>(Corresponding Author) Professor, School of Management and Economics, Handong Global University, Handongro 558, 37554, Korea, Email: jooheeh@handong.edu

## TVAE를 이용한 영 과잉 데이터 생성 연구

길경주<sup>1</sup>, 전종준<sup>2</sup>

### 요 약

신뢰도 높은 AI시스템을 구축하기 위해서 양질의 데이터가 필요하지만 AI모델의 훈련에 필요한 데이터를 구하기는 쉽지 않다. 원본 데이터의 통계적 특성을 모방한 합성 데이터(재현 데이터)는 고질적인 데이터 병목현상을 해소할 수 있다. 합성 데이터를 만드는 방법론으로는 딥러닝 모델이 빈번하게 사용되고 있다. 본 연구에서는 테이블 자료(Tabular Data)에 적용 가능한 TVAE 모델을 사용하여 금융데이터 재현자료 생성 방법을 제안한다. 특히, 금융 데이터에서 빈번하게 발생하는 영 과잉 데이터를 원 데이터와 유사한 비율로 재현해내는 것을 목표로 한다. 기존 TVAE 모델을 활용해 영 과잉 컬럼을 재현해낸 결과 영인 데이터의 비율이 원 데이터와 비교하여 현저히 떨어지는 것을 알 수 있었다. 따라서, 해당 모델에 이진분류기를 도입하여 영인 데이터의 비율을 추정하고, 이를 바탕으로 원 데이터와 유사한 비율을 가진 영 과잉 데이터를 재현해내고자 하였다. 또한, 클래스 불균형이 있어 희소한 값을 가지는 클래스를 잘 재현해내지 못하는 범주형 컬럼에 대하여 Gumbel-Softmax Trick을 적용하여 원 데이터의 분포와 유사한 재현 데이터를 생성해낼 수 있었다. 재현 데이터의 평가지표로는 머신러닝 유용성(Machine learning utility)과 히스토그램, 막대그래프와 같은 시각화 자료를 이용한다. 재현 데이터의 보안상 안전지표로는 DCR과 Attribute Disclosure를 사용한다.

주요용어 : 생성모델, 변이형 오토인코더, 영 과잉 모델, 테이블 데이터.

<sup>1</sup>02504 서울특별시 동대문구 서울시립대로 163(전농동)90, 서울시립대학교 통계학과 석사과정.  
E-mail: qw4735@uos.ac.kr

<sup>2</sup>(교신저자)02504 서울특별시 동대문구 서울시립대로 163(전농동)90, 서울시립대학교 통계학과 교수.  
E-mail: jj.jeon@uos.ac.kr

## 역확률 가중치 추정량과 그래프 라쏘를 이용한 고차원 자료에서 다중대체를 위한 변수선택방법 제안

이윤아<sup>1</sup>, 박성오<sup>2</sup>

### 요약

생물 의학이나 사회조사 분야의 자주 발견되는 결측자료에 대하여 결측치를 모두 제거하고 분석을 수행하면 정보의 손실과 함께 추정량에 편향이 생길 수 있다. 이를 해결하기 위해 각 결측값에 대하여 여러개의 ‘그럴듯한’ 값을 생성하여 대체한 후 분석결과들을 재결합하는 다중대체법(multiple imputation; MI)이 흔히 사용되어 왔다. 다중 변수가 결측값을 갖는 일반적인 결측자료 패턴에서 흔히 사용되는 MICE (multiple imputation by chained equations) 방법은 결측이 포함된 각 변수를 나머지 변수들로 예측하는 모형을 학습하여 결측을 메우는 과정을 반복하게 되므로, 고차원 자료의 경우 변수선택을 고려하여 모형을 학습하는 것이 필요하다. 본 연구에서는 MICE 알고리즘의 각 단계에서 사용하는 변수 집합을 추정하기 위해 변수 간의 조건부 독립성을 표현하는 역공분산 행렬을 이용한 새로운 변수선택방법을 제안한다. 성긴(sparse) 역공분산 행렬을 추정하기 위해 그래프 라쏘(graphical LASSO) 모형을 사용하며, 이때 사용되는 공분산 행렬 추정량으로 결측치에 대한 보정을 고려한 역확률 가중치(inverse probability weighting; IPW) 추정량을 사용하고자 한다. 모의실험을 통해 기존의 두 가지 고차원 MI를 위한 변수선택 방법들 - MICE 시 벌점화 회귀를 직접적으로 이용하는 DURR (direct use of regularized regression), 간접적으로 이용하는 IURR (indirect use of regularized regression) - 과의 성능을 비교한다.

**Keywords** : 결측치, 다중대체법, 고차원자료, 변수선택, 그래프라쏘, 역확률가중치.

### References

- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. "Sparse inverse covariance estimation with the graphical lasso." *Biostatistics* 9.3 (2008): 432-441.
- Deng, Yi, et al. "Multiple imputation for general missing data patterns in the presence of high-dimensional data." *Scientific reports* 6.1 (2016): 21689.

<sup>1</sup>02844 서울 성북구 보문로34다길 2, 성신여자대학교 통계학과 석사과정. E-mail: dbsdk3337@naver.com

<sup>2</sup>02844 서울 성북구 보문로34다길 2, 성신여자대학교 통계학과 조교수. E-mail: spark6@sungshin.ac.kr

## 소상공인 자립을 위한 서울시 행정동별 폐업 리스크 예측 모델

송예진<sup>1</sup>, 박지호<sup>2</sup>, 백예은<sup>2</sup>, 최정빈<sup>2</sup>, 김동하<sup>1,2</sup>

### 요약

현재 증가하는 개업 상권의 규모만큼 폐업하는 상권 또한 증가하고 있다. 이에 따른 정교한 상권분석이 필요한 시점이다. 현재의 상권분석 시스템들은 단순한 통계 수치만을 보일 뿐 리스크 제공은 하지 않고 있다. 이를 해결하기 위해 본 연구에서는 소상공인 상점 개업 시 폐업 리스크를 예측하는 모형을 개발한다. 이를 위해 개·폐업 데이터(마이 데이터)를 기반으로 상권 활성화 지수, 생활인구, 행정동별 추정 매출, 임대 시세 데이터를 복합하여 개인화 폐업 리스크를 예측하는 부스팅 알고리즘 기반의 머신러닝 모델을 개발한다. 해당 모형은 서울시 행정동에서부터 추출된 개·폐업 상점에 대한 폐업 여부를 예측한 결과 AUC 0.61의 성능을 보였다. 본 연구는 업종, 개업 위치별 폐업 리스크 도출 및 업종별 최적의 위치 추천, 행정동별 최적의 개업 업종 추천이 가능하다는 장점이 있다. 따라서 소상공인이나 투자자에게 중요한 비즈니스 결정 및 리스크 관리를 위한 정보를 제공하고, 새로운 경제 지원 정책을 이끌어내 폐업률을 낮추는 데 기여할 수 있을 것이다.

주요용어 : 폐업 리스크 예측, 최적 업종 및 위치 추천, 부스팅 알고리즘.

<sup>1</sup>성신여자대학교 통계학과.

<sup>2</sup>성신여자대학교 수리통계데이터사이언스학부.

## DBSCAN을 활용한 구성자료 군집분석\*

이제석<sup>1</sup>, 김병원<sup>2</sup>

### 요약

본 연구는 구성자료(Compositional Data)의 군집분석에 있어 거리 계산 방법에 따른 군집의 분포 및 특성에 차이가 발생하는지를 파악하기 위해 밀도 기반 군집분석 방법인 DBSCAN(Density-Based Spatial Clustering of Applications with Noise) 알고리즘을 활용한다. 구성자료는 구성요소가 갖는 값의 총합이 같은 자료를 의미하며, 표본공간이 심플렉스(simplex)로 정의되어 일반적인 벡터공간에서 정의된 거리 계산 방법을 적용할 수 없다. DBSCAN 알고리즘은 데이터의 크기가 크고 불규칙한 형태를 보이는 공간자료에 대해서도 밀도 기반의 계산방식으로 유연하고 강력한 성능을 보이는 군집분석 방법으로 다양한 형태의 자료를 분석하는데 유용한 방법이다. 본 연구에서는 구성자료 간의 거리 계산 방법으로 대표적으로 사용되는 Aitchison 거리와 이 거리 계산 방법의 단점을 보완할 수 있는 한 가지 대안인 방향거리(directional distance)를 사용하여 군집분석을 시행한 결과를 비교 분석한다. 구성자료의 다양한 분포 형태에 따라 두 거리 계산 방법에 어떠한 차이가 있는지를 파악하기 위해 dirichlet 분포를 활용한 모의 실험을 진행하였다. 또한 계산 방법 별 군집분석 결과에 차이가 있는지를 확인하기 위한 실제 예시로, 미국프로농구(NBA)의 공격 시 포지션 별 볼 점유율에 해당하는 USG%(Usage Percentage)를 활용하여 93-94시즌부터 21-22시즌까지의 동, 서부 컨퍼런스 1위 팀들을 대상으로 군집분석을 시행한 결과를 비교하였다.

주요용어 : 구성자료, 군집분석, DBSCAN, 방향거리, Aitchison거리.

\*이 논문은 2022년도 정부 (교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구 사업임 (한국연구재단에서 부여한 과제번호: NRF-2021R1G1A109535511-1-1).

<sup>1</sup>41566 대구광역시 북구 대학로 80, 경북대학교 통계학과, 석사과정. E-mail: dlakakwns@knu.ac.kr

<sup>2</sup>(교신저자) 41566 대구광역시 북구 대학로 80, 경북대학교 통계학과, 교수. E-mail: byungwonkim@knu.ac.kr

## Box monitoring of mixed-type multivariate data using Gower-type distance

*Woohyun Kim<sup>1</sup>, Sungim Lee<sup>1</sup>, Johan Lim<sup>2</sup>*

### Abstract

The analysis and monitoring of mixed-type multivariate data, consisting of continuous, categorical, and count variables, poses significant challenges in various fields. This study focuses on applying the Gower-type distance, which considers the heterogeneity of variable types and enables comprehensive analysis by capturing the data's quantitative and qualitative aspects. This research addresses the challenges of analyzing and monitoring mixed-type multivariate data using the Gower-type distance and Hotelling's chart. Our method involves comparing new mixed-type observations with historical data using the Gower-type distance and proposing a novel Hotelling's chart based on box statistics derived from the quartiles of the Gower-type distance. To evaluate the performance of our method, we perform extensive simulation studies comparing it to an existing method. Through these studies, we assess the effectiveness of our approach in detecting anomalies and maintaining control in different scenarios. Finally, we demonstrate our proposed method's practical applicability and ability to handle mixed-type multivariate data using real-world data.

*Keywords* : Mixed-type data; Gower-type distance; Multivariate control chart; Phase-II monitoring.

---

<sup>1</sup>Department of Statistics, Dankook University 152 Jukjeon-ro, Suji-gu, Yongin-si, Gyeonggi-do, 16890, South Korea. E-mail: wooh0514@dankook.ac.k

<sup>2</sup>Department of Statistics, Seoul National University.

## LDA를 이용한 HMD 기반 가상 현실 장치 사용자 경험 분석 연구\*

유일선<sup>1</sup>

### 요 약

HMD (Head-Mounted Display) 기반 가상현실은 사용자에게 고도의 존재감과 몰입감을 제공할 수 있는 가상현실 시스템 중 하나로, 이는 가상현실 환경에서의 사용자경험 (UX: User eXperience)의 핵심 가치이다. 또한, 사용자 경험을 관찰하는 것은 사용자의 내재적인 요구를 파악하는 데 있어서 중요하다. 그러나, 이전 연구들은 주로 실험실 환경에서 자연스럽게 않은 사용 상황에 대한 관찰을 수행해왔기 때문에 자연스러운 사용 맥락에서 사용자의 행동을 관찰하는 것이 어려웠다. 온라인 리뷰 (예: amazon.com 등)를 크롤링하여 사용자의 말을 통해 다양하고 자연스러운 UX를 관찰하는 데 도움이 될 수 있다. 이는 수백만 사람들의 생각에 관한 정보를 파고 들기 위해 활용될 수 있는 잠재적으로 가치 있는 데이터 소스이다. 따라서, 본 연구의 목적은 HMD를 기반으로 한 가상현실 시스템의 사용자 경험을 텍스트 마이닝 기법과 토픽 모델링 기법을 활용하여 파악하는 것이다. 본 연구에서는 HMD 기반 가상현실 장치의 사용자 경험을 수집하기 위해 외부 데이터를 텍스트 마이닝하는 기법을 사용하였다. 수집된 외부 데이터에서는 토픽 모델링 기법 중 LDA (Latent Dirichlet Allocation)을 이용하여 HMD 기반 가상현실 장치의 사용자 경험의 주요 이슈나 주제를 파악하였다. 분석 결과, 총 19,288개의 HMD 기반 VR 장치 리뷰가 amazon.com에서 수집되었다. 결과는 LDA 기법을 통해 HMD 기반 가상현실 장치의 사용자 경험에 대한 다섯 가지 주제 (토픽 1: 기능 및 스펙, 토픽 2: Presence, 토픽 3: 사이버 멀미, 토픽 4: 사용성, 토픽 5: 재미 및 엔터테인먼트)가 도출되었다. 본 연구는 텍스트마이닝 기법과 토픽모델링 기법을 사용하여 편향 없이 사용자 경험을 효과적으로 정량적으로 수집하고 분석할 수 있다는 것을 제안하였다. 이를 통해 제품 및 서비스에 대한 사용자의 내재적인 요구를 이해하는 데 도움이 될 것으로 기대된다.

주요용어 : 토픽모델링, 텍스트마이닝, LDA (Latent Dirichlet Allocation), 사용자경험, 가상현실

\*이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1F1A1075846).

<sup>1</sup>(교신저자) 02748 서울시 성북구 화랑로13길 60, 동덕여자대학교 HCI사이언스전공 조교수.  
E-mail: isrhiu@dongduk.ac.kr

## 양상블 비모수적 변수 선택 방법 : 워게임 데이터 분석\*

신웅섭<sup>1</sup>, 차영호<sup>2</sup>, 한봉규<sup>3</sup>

### 요약

빅데이터의 등장은 우리의 삶과 산업 전반에 큰 영향을 주고 있다. 단순히 데이터를 수집하고 저장하는 수준에서 벗어나, 데이터를 분석하고 이를 이용하여 새로운 가치를 창출하는 기술의 발전을 가져오고 있다. 이러한 변화의 흐름은 군이 보유하고 있는 데이터에 대한 인식과 태도에도 영향을 미치고 있다. 군 데이터의 특성상 보안이 강조되다 보니 생성된 데이터를 안전하게 저장, 관리하는 수준에서 크게 벗어나지 못했던 과거와 달리, 최근에는 군에서 축적해오던 다양한 데이터를 통계 기법들을 활용하여 분석하려는 노력들이 활발하게 이루어지고 있다. 특히, 과학화 전투훈련단(Korea Combat Training Center, KCTC)이 보유한 각 전투원들에 대한 훈련 데이터와 전투지휘훈련단(Center for Battle Command Training Program, BCTP)에서의 워게임 데이터들은 전장상황을 간접적으로 체험한 결과물로서 미래 전장상황 예측을 위한 핵심적인 데이터로 활용되고 있다. 본 논문에서는 양상블 비모수적 변수 선택 방법(Shin, 2023)을 BCTP 워게임 데이터에 적용하여 실제 교리적, 전술적으로 타당한 결론 도출이 가능한지 확인한다.

주요용어 : 양상블 기법, 변수중요도, 유의성, 워게임 데이터.

\* 이 논문은 제1저자 신웅섭의 석사학위논문(Shin, 2023)을 바탕으로 추가 연구하여 작성한 것입니다.

<sup>1</sup>34059 대전광역시 유성구 자운로 97번길 사서함 78-5호, 육군 전투지휘훈련단 기동논리장교.

E-mail: triple0112@naver.com

<sup>2</sup>34059 (교신저자) 대전광역시 유성구 자운로 97번길 사서함 78-5호, 육군 전투지휘훈련단 전투모의과장.

E-mail: infcha@empas.com

<sup>3</sup>34059 대전광역시 유성구 자운로 97번길 사서함 78-5호, 육군 전투지휘훈련단 전투모의처장.

E-mail: polarishan@hanmail.net



## 중국 대학생의 연애편 분석

평인란<sup>1</sup>, 양정호<sup>2</sup>

### 요약

연애는 대부분의 개인이 평생 동안 직면해야 하는 주제이며 특히 성인 초기 개인에게 특히 중요하다. 미국의 정신분석학자 에릭슨은 누구나 성장하면서 보편적으로 생물적, 생리적, 사회적 사건의 발달순서를 경험하고 단계별로 발전한다고 보고, 성년의 초기단계에 있는 개체(18-25세)의 발달임무는 친밀감을 얻고 외로움을 피하며 사랑의 실현을 경험하는 것이라고 보았다. 대학생의 심리적, 생리적 발달은 나날이 성숙해지고 있으며, 학습 및 생활환경의 상대적 개방성과 자유로움, 대학생의 충분한 시간과 공간적 발달과 이성교제, 대학생의 연애는 자유화, 보편화 및 공개적인 집단현상이 되었다. 뿐만 아니라, 시대의 흐름에 따라 청소년들의 연애도 보편화되고 있으며, 연애 연령은 여전히 젊어지고 있다는 조사도 있다. 중국에서는 고등학교 졸업 전까지 연애를 절대 금지하는 일도 있고 학교와 학부모는 연애를 하면 공부에 지장이 있다는 이유를 드는 경향이 있다. 대학에 입학한 청소년들은 학부모와 선생님의 울타리에서 벗어나 마침내 자기하고 싶은 일을 자유롭게 할 수 있게 된 것이다. 따라서 본 연구에서는 현재 중국 대학생들의 연애심리와 연애편의 특징을 파악하여 그들이 올바른 연애편을 확립할 수 있는 방법을 찾아보려고 한다. 이를 위해 중국내 대학생 153명을 대상으로 대학생들의 연애심리 및 연애편 등에 대해 온라인 설문조사를 실시하여 그 조사결과를 분석하려고 한다. 분석결과로는 대부분의 대학생은 연애 문제에 대해 긍정적이거나 크게 상관을 하지 않는 것으로 나타났다. 대학생의 연애가치관의 문제 중, 대학생은 배우자 선택에 있어 대부분 인품, 재능, 성격이 중요한 선택기준으로 나타났다. 실연을 당했을 때 다수의 학생이 받아들일 수 있다고 답했지만 여전히 15.1%가 받아들일 수 없다고 답했고, 3.5%는 자살생각까지 든다고 답했고, 1 0.9%는 확실치 않다고 답했으며, 8%는 보복 심리를 가지고 있었다. 따라서 현재 대학생들의 연애는 이미 공식화되고 보편화되고 있지만, 일부 대학생들의 연애편에 대한 연구는 연애동기와 배우자 선택기준, 연애좌절능력 및 대처방식, 연애와 학업의 균형등이 낙관적이지 못하다는 것을 밝혀냈기 때문에 이에 따라 대학생들에게 연애 관련 심리상담교육을 실시할 필요가 있다.

주요용어 : 중국 대학생, 연애편, 설문조사, 연애편 교육.

<sup>1</sup>03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: fengyinran@icloud.com  
<sup>2</sup>(교신저자) 03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

## Analysis of Chinese high school students' motivation to study abroad and its external factors

*Jin Hua*<sup>1</sup>, *Jung-Ho Yang*<sup>2</sup>

### Abstract

Under the influence of the increasing popularity of self-funded study abroad in China and the younger age of studying abroad, the proportion of high school students who choose to study abroad is increasing rapidly. This study randomly selected 7 students in a high school in Beijing who were preparing to study abroad, and conducted in-depth interviews with them. The results show that there are 3 motivations for high school students to study abroad at their own expense, 1) Reduce the pressure of the Chinese college entrance examination 2) Prefer foreign education and lifestyle 3) Be helpful for employment. Moreover, according to the content of the interview, it can be found that the following three external reasons also have an impact on high school students' study abroad: 1) The huge profits of study abroad agencies drive more people to join the industry, which indirectly stimulates the growth of the number of study abroad. 2) Chinese parents are paying more and more attention to their children's education, and more parents are hesitating to increase their education budget in order to allow their children to receive a better education. 3) Foreign schools are affected by the the shortage of the population, resulting in the problem of insufficient students, in order to maintain the operation of the school. Thus, they are willing to absorb more international students, China, as the main source of international students, is also affected by it.

*Keywords* : self-funded study abroad, high school student, motivation to study abroad, educational agents.

---

<sup>1</sup>Graduate Student, Department of the Education, Sungkyunkwan University, 25-2, Sungkyunkwan-Ro, Jongno-gu, Seoul, 03063, Korea. E-mail: jinhua9811@163.com

<sup>2</sup>Professor, Department of the Education, Sungkyunkwan University, 25-2, Sungkyunkwan-Ro, Jongno-gu, Seoul, 03063, Korea. E-mail: jyang@skku.edu

## 중국 대학의 국제화 지표 및 순위 분석

왕호<sup>1</sup>, 양정호<sup>2</sup>

### 요약

최근 대학교육이 보편화되면서, 전세계적으로 대학을 중심으로 지식 경쟁과 인재 경쟁을 핵심으로 하는 경쟁이 치열해지고 있다. 인재 경쟁은 주로 대학교육에 상당부분 의존하고 대학은 핵심 우수인재를 양성에 집중하면서 대학교육의 질이 무엇보다 중요하다. 특히 국가간 교류가 확대되면서, 대학교육의 국제화를 통해 자국의 문화와 가치관을 외국대학과 잘 교류하고, 국제적인 영향력을 높이면서 대학마다 국제화에 많은 노력을 기울이고 있다. 본 연구는 2023년 상하이교통대학교가 전세계 대학을 중심으로 평가를 진행한 대학평가순위(Shanghai Ranking's Academic Ranking of World Universities, 약칭 ARWU) 자료를 이용하여 중국 대학의 국제화 지표와 특성을 분석하려고 한다. 수집된 데이터에서 특히 중국 대학의 다양한 국제화 지표를 살펴보고, 지표활용시 나타나는 현재의 어려움도 분석하려고 한다. 주요 연구 결과는, 첫째, 일반 대학교에 비해 중국 최고의 대학 상위 10개 대학교가 국제화 지표 중 유학생 확보가 급증하는 모습을 보이고 있다. 둘째, 유학생 분포 상황을 보면 중국 동부 지역의 대학에는 유학생이 더 집중하고 있다. 중국에 온 유학생 수 상위 10개 성(省) 중 후베이성만 중부에 위치하고 기타 9개 성(省)은 모두 동부 지역에 속하는 결과를 나타냈다. 셋째, ARWU 대학 순위 체계로 보면 중국 100위권 대학의 국제화 평가지표 점수 상위 대학과 하위 대학 간 격차가 있고 중국 대학교의 학술 수준은 해마다 향상되고 있지만 국제화 지표나 수준은 세계 주요대학과 비교해 큰 차이가 존재하고 있다.

주요용어 : 중국 대학교육, 국제화 지표, ARWU 세계 대학 순위, 대학 국제화 문제.

<sup>1</sup>03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: 985043909@qq.com

<sup>2</sup>(교신저자) 03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

## 중국 교원순환 기준 데이터 분석

엽혜림<sup>1</sup>, 양정호<sup>2</sup>

### 요약

최근 중국 정부는 교육이 국가 미래의 원동력이자 사회 안정화에 중요한 역할을 한다고 보고 교육 개혁을 단행했다. 교육 개혁의 일환으로 2021년 7월에 지역의 학교마다 교사를 전보형태로 순환근무하게 하는 교원순환제 시행을 예고했다. 교사 질을 비롯해 아직 지역마다 교육격차가 심각한 중국의 교육 자원 불균형 문제는 오래된 숙제이다. 최근 몇 년 동안 ‘학교선택’, ‘명문학군의 집값 상승’ 등의 사회 문제는 교사를 비롯한 교육 자원의 불균형을 심화시켰다. 중국 교육당국은 교육 자원의 불균형 문제를 해결하기 위해 다양한 노력을 기울인지 10년이 넘었지만 지금까지 제대로 시행되지 않고 관심도 부족한 교원순환제를 학교가 교육격차 해소위해 대대적으로 추진하고 있다. 이전과 달리 이번에 교원순환제는 큰 주목을 받으면서 교육 개혁에 강한 의지를 밝혔다. 본 연구로 중국의 교육 개혁의 중요한 조치인 교원순환제 시행과정에 나타난 문제점을 탐색하고, 교원순환 기준이 되는 데이터를 분석해 향후 중국 교원순환제 정책 발전에 기여할 수 있기를 기대하고 있다. 본문은 문헌연구법을 채택하여 중국 정부가 발표한 공식 문건, 선행 연구 자료, 각종 중국 국가 회의 자료, 교원순환제에 관한 기사 등에서 교원순환제 시행을 위한 교원의 순환기준 자료를 비교분석하였다. 본 연구는 먼저 중국 교원순환제 정책의 시행 배경과 내용을 알아보고, 의무교육단계의 교원순환제 시행 과정에 존재하는 문제점을 분석했다. 또 교원순환제 시행을 위한 교원 근무연수, 거주지역 등의 관련 순환기준의 데이터간 차이를 살펴 보면서, 교원순환제에 교원 소극적 대응, 정책의 형식적 시행, 안정성 결여, 정책체계 부실 등의 문제점을 살펴보았다. 마지막으로 이런 문제점을 해결하기 위해 교원순환제의 안정적 정착을 위한 대안도 제시하였다.

주요용어 : 중국 교원순환제, 교원순환 기준자료, 학교간 교사 질 격차, 의무교육확대.

<sup>1</sup>03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: yhl1277713099@gmail.com  
<sup>2</sup>(교신저자) 03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

## 중국 가오카오이민 관련 입학자료 분석

손옥하<sup>1</sup>, 양정호<sup>2</sup>

### 요약

‘가오카오이민’은 중국의 복잡한 사회 환경에서 비롯된 일종의 교육현상으로, 명문대학에 진학하기 위해 상대적으로 대입시험에서 조금이라도 유리한 지역에서 시험을 보려는 ‘이민현상’이다. 이런 문제를 제대로 관리하지 않으면 교육의 형평성과 사회 풍토 등에 악영향을 미칠 수 있다. 이 문제를 해결하려면 여러 분야에서 종합적인 분석이 필요하지만, 이 연구에서는 이 현상의 이론적 내용을 파악하고 ‘가오카오이민’의 실제 현황을 입학자료를 이용해 분석해서 교육의 공정성을 유지하고 향후 교육제도 개혁에 참고할 수 있도록 하려고 한다. 본 연구는 중국 각 성의 대학 자원 분포 현황, 대학 입학시험인 가오카오 합격선 차이, 기초교육 현황 등의 데이터를 통해 ‘가오카오이민’ 현상의 원인을 분석하였다. 주요 연구 결과에 따르면 첫째, 중국의 대학은 지역별로 상당히 불균등하게 분포되어 있으며 명문대학은 주로 동부 경제 발전 지역에 집중되어 있고 중서부의 교육 수준은 상대적으로 낮아져 있다. 둘째, 중국 각 성의 대입시험인 가오카오 합격선과 합격률에는 큰 차이가 있다. 셋째, 중국 각 지역의 기초교육 발전의 불균형은 각 지역의 교육재정예산과 교사의 학력 차이로 나타난다. 본 연구에서는 ‘가오카오이민’ 현상에 대한 탐구를 더욱 개선하기 위해 자료분석에 더해 6명의 가오카오이민자를 대상으로 인터뷰를 진행하는 질적 연구도 수행하였다. 이를 통해 ‘가오카오이민’의 원인과 실제사례 분석을 하였으며, ‘가오카오이민’의 예방 및 해결방향을 제시하였다.

주요용어 : 가오카오이민, 중국 대입시험 분석, 가오카오 입시합격 격차, 명문대학 분포 비교.

<sup>1</sup>03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 석사과정. E-mail: sunyuxiakr@163.com

<sup>2</sup>(교신저자) 03063 서울시 종로구 성균관로 25-2, 성균관대학교 교육학과 교수. E-mail: jyang@skku.edu

## 원거리 시각 촛점 고정과 어지러움증과의 관계

노효련<sup>1</sup>, 홍성현<sup>2</sup>, 전상준<sup>3</sup>, 안소현<sup>3</sup>, 구교준<sup>3</sup>, 이유진<sup>3</sup>, 이수빈<sup>3</sup>, 이영희<sup>4</sup>

### 요약

본 연구는 시각 고정 유무 후 어지러움증으로 인한 균형과 전정 감각 정도를 알아보려고 하였다. 시각 고정 어지러움증을 유발하기 위하여 바로 선 자세에서 10바퀴를 돌은 후에 전정감각, 균형, 보행 검사를 각각 실시하였다. 연구 대상자는 20대 성인 38명(남: 21명, 여:17명)이었다. 전정감각 정도를 측정하기 위하여 8자 모양 경로 보행 검사를 실시하였고 균형 능력 검사는 한다리 들고 서기 검사, 보행과 균형 능력 정도는 Timed up & go 검사를 실시하였다. 균형과 전정감각 및 보행능력은 성별과 시각 고정 유무에서 차이가 나타나지 않았다.

주요용어 : 어지러움증, 시각, 보행, 균형, 전정감각.

### 1. 서론

멀미는 동요병(motion sickness)라고도 하며, 비행기, 바다, 육지, 우주 여행자들에게 일시적으로 나타나는 질병(illness)이다. 전 인구의 5%는 심한 멀미를 경험하게 되며, 나머지는 가벼운 멀미를 경험한다고 한다(Han & Jang, 2008). 일반적으로 멀미는 전정감각에 의하여 나타난 이상 증상을 의미하는 경우가 많다. 하지만 멀미는 시각으로만 발생할 수도 있고, 시각과 전정감각 간의 감각 부조화로 발생할 수도 있다. 따라서 멀미는 시각과 전정 감각 간의 불일치에 의해 발생하는 신체 반응 전체를 포함한다.

이 연구에서는 멀미의 증상 중 하나인 어지러움증이 시각과 전정감각과의 관계를 알아보려고 한다.

### 2. 연구 방법

연구 대상자는 20대 성인 38명(남: 21명, 여: 17명)이었다. 연구 대상자들은 1차와 2차 실험을 진행하였는데 우선, 어지러움증을 유발하기 위하여 바로 선 자세에서 10바퀴를 돌은 후에 전정감각, 균형, 보행 검사를 각각 실시하였다. 2차 실험에서는 1m 정도 떨어진 특정 지점을 보고 바로 선 자세에서 10바퀴를 돌은 후에 전정 감각, 균형, 보행 검사를 각각 실시하였다.

<sup>1</sup>25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 교수. E-mail: withry@kangwon.ac.kr

<sup>2</sup>(교신저자) 25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 학부재학생.  
E-mail: j000154@naver.com

<sup>3</sup>25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 학부 재학생.

<sup>4</sup>48015 부산시 해운대구 반송순환로142, 영산대학교 미용건강관리학과 부교수. E-mail: yhlee@ysu.ac.kr

전정감각 정도를 측정하기 위하여 8자 모양 경로 보행 검사(Figure-of-Eight Walk Test, F8WT)를 실시하였고 균형 능력 검사는 한다리 들고 서기 검사(one leg test), 보행과 균형 능력 정도는 Timed up & go(TUG) 검사를 실시하였다. 8자 모양 보행검사는 전정감각을 평가하는 검사법으로 8자 모양 트랙은 2개의 장애물 주변을 8자 모양으로 보행하는 검사로, 2개의 장애물 중간에서 시작하여 반 시계 방향으로 왼쪽의 장애물을 돌아 시작 지점으로 왔다가 시계 방향으로 나머지 장애물을 돌아오는 시간을 측정한다. 본 연구에서는 출발점에서 첫 스텝이 시작하는 순간부터 두 발이 제자리로 돌아와 종료되는 시점까지를 초시계로 측정하였다. 한다리 들고 서기 검사는 한 다리로 1분 동안 서서 유지할 수 있는지를 알아보는 검사이며 TUG 검사는 앉은 자세에서 신호에 따라 일어나서 3m 떨어진 곳의 반환점까지 가서 돌아와서 앉기까지의 시간 측정하며 이 연구에서는 두발을 일직선으로 유지하면서 걷는 시간을 측정하였다.

수집된 자료는 통계 프로그램 SPSS 21.0을 이용하였으며, 독립표본 t-test와 대응표본 t-test를 실시하였다. 유의수준은  $p < .05$ 로 설정하였다.

### 3. 연구 결과

#### 1) 성별과 전정감각, 균형, 보행정도

성별에 따른 전정감각과 균형, 보행정도를 알아보았다. 전정감각, 균형, 보행정도는 성별에 따른 통계학적인 차이가 나타나지 않았다( $p > .05$ ). 또한, 1m 떨어진 곳을 고정 한 후의 전정감각, 균형, 보행정도는 성별에 따른 통계학적인 차이가 나타나지 않았다( $p > .05$ ).

Table 1. Degree of balance and vestibular sense

Symptom	Gender	Mean	SD	t	p
One leg	M	11.12	11.07	-1.445	0.157
	F	16.78	13.06		
F8WT	M	7.25	1.27	-0.379	0.707
	F	7.39	0.96		
TUG	M	16.76	4.94	-1.196	0.239
	F	18.72	5.14		
Fixed focus oneLeg	M	13.46	12.31	0.406	0.687
	F	11.82	12.38		
Fixed focus F8WT	M	7.22	1.19	-0.517	0.608
	F	7.41	0.99		
Fixed focus TUG	M	16.90	5.87	-0.633	0.531
	F	17.99	4.37		

#### 2) 사이버 멀미 증상과 가상현실 체험횟수 및 멀미 정도와의 상관관계

시각 고정 없이 어지러움증을 유발한 경우와 1m 떨어진 곳을 고정 한 경우에서의 전정감각과 균형, 보행정도를 알아보았다. 전정감각, 균형, 보행정도는 통계학적인 차이가 나타나지 않았다( $p > .05$ ).

Table 2. Fixed focus and balance, vestibular sense

	Mean	SD	t	p
OneLeg	13.65	12.17	0.459	0.649
Fixed focus oneLeg	12.73	12.20		
F8WT	7.31	1.13	0.039	0.969
Fixed focusF8WT	7.30	1.09		
TUG	17.64	5.06	0.777	0.442
Fixed focusTUG	17.39	5.22		

#### 4. 결론

멀미는 비행기, 바다, 육지, 우주여행 등 모든 교통수단에서 경험할 수 있으며, 가상현실 체험시 나타나는 사이버 멀미(cybersickness)도 관심을 받고 있다. 멀미에 대한 원인은 확실하게 밝혀지지 않았으나, 전정기관, 시각, 그리고 근육, 다른 조직의 고유수용체에서 받은 정보가 서로 달라 뇌에서 충돌을 일으켜서 나타난다는 “conflict theory”가 전통적인 이론이다.

본 연구에서는 원거리의 시각 고정과 시각 고정이 없는 경우에서의 어지러움증을 유발시킨 후 전정감각과 균형에 어떤 영향이 있는지 알아본 결과 그 차이가 나타나지 않았다. 즉 시각 고정유무는 전정감각이나 균형에 영향을 주지 않는 것으로 나타났다. 본 연구는 제자리 돌기로 어지러움증을 유발하여서 실제의 멀미 상황보다 그 강도가 낮고 시간이 짧아서 이 연구결과를 멀미의 상황에 적용하는 것은 무리가 있는 것으로 보인다.

#### References

- Han, Y. J., Jang, G. T.(2008). The Cause of Motion Sickness in Oriental Medicine. *J Korean Oriental Pediatrics*, April, 2008;22(1):149-162.
- Cho, H. C., Roh, H. L., Lee, H. J. (2022). The Effect of Convergence Vision Therapy on VR Cybersickness, *Journal of The Korea Convergence Society*, 13(2). 55-65.



## 국내·외 악력 측정기의 악력 차이 분석

노효련<sup>1</sup> 김다정<sup>2</sup>, 이한민<sup>2</sup>, 이지은<sup>2</sup>, 임철균<sup>2</sup>, 박윤아<sup>2</sup>, 김혜미<sup>2</sup>, 이영희<sup>3</sup>, 이효택<sup>4</sup>

### 요 약

손 근력은 활용성이 높은 지표로 이나 측정이 용이하다는 장점을 가지고 있다. 본 연구에서는 국내 악력측정기와 중국, 일본 악력 측정기 간의 악력 정도를 비교하여 그 차이를 알아보고자 하였다. 연구 대상자는 건강한 20대 성인 49명(여자 28명, 남자 21명)이었고 세 종류의 악력 측정기를 이용하여 각각의 악력을 측정하였다. 측정된 악력 정도는 차이가 있었는데, 특히 한국제품이 일본과 중국과 차이가 나는 것으로 보인다. 한국제품에서 여자의 악력은 가장 높게 나왔고 남자는 가장 약하게 나타났다. 따라서, 한국산 악력 측정기의 악력의 정도에 대한 민감도가 떨어지는 것으로 나타나서 보완이 필요하다고 여겨진다.

주요용어 : 악력, 악력 측정, 한국, 일본, 중국.

### 1. 서론

인체는 노화함에 따라 근육량과 근력이 점점 감소하는데, 악력은 측정하기 쉬우면서도 현재의 전반적인 근력과 영양 상태, 근육량을 평가할 수 있는 매우 효율적인 지표이다. 악력은 근 감소증을 진단하는 임상적인 지표로써도 활용되고 있는 측면이 있으며, 국민의 건강 및 영양상태를 정확히 파악하여 국가건강정책을 수립하기 위해 질병관리청에서 수행하는 법정 조사인 국민건강 영양조사에서도 악력은 사용되고 있다.

또한, 제2형 당뇨병 환자에서 낮은 악력을 나타내고 있어서 당뇨에 대한 예측 인자로 활용될 수도 있다(윤은선과 박수현, 2020; 홍성현 등, 2021). 악력은 노인들의 인지장애와 치매 정도를 예측하는데에도 활용될 수 있는데, 건강한 노인보다 인지 장애를 가진 노인에서 손 근력과 민첩성이 유의하게 낮다(Kimi et al, 2017).

손 근력은 활용성이 매우 높으며 측정이 용이하고 간편한 정보이나 신뢰도있는 국내 손 근력 측정기가 부족한 상황이다. 이에 본 연구에서는 국외 제품과 국내산 악력 측정기 간의 차이를 알아보고자 한다. 이를 통해 국내산 제품의 기술력을 확인하고 활용도를 높이고자 한다.

<sup>1</sup>25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 교수. E-mail: withtry@kangwon.ac.kr

<sup>2</sup>25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 학생. E-mail: dahong20100@gmail.com

<sup>3</sup>48015 부산시 해운대구 반송순환로142, 영산대학교 미용건강관리학과 부교수. E-mail: yhlee@ysu.ac.kr

<sup>4</sup>(교신저자) 26462 강원도 원주시 황금로8 ㈜바이오헬스케어 대표. E-mail: hyotaek@biohealthcore.com

## 2. 연구 방법

### 1) 연구 절차

국내외 악력 측정기 간의 신뢰도를 평가하기 위하여 중국산(Camry, China), 일본산(Takai, Japan), 한국산(Soundbody-grip, Korea)의 세종류를 사용하였다. 연구 대상자는 20대 학생 49명(남자:21명, 여자:28명)을 대상으로 하였으며, 이들은 상지 질환이 없는 건강한 성인들이었다. 모든 연구 대상자들의 손길이와 악력 정도를 세종류의 악력 측정기를 이용하여 각각 2회씩 측정하였다.

### 2) 분석 방법

측정한 악력 정도는 측정한 두 번의 악력 평균값을 사용하여 분석하였다. 일반적 특성은 기술통계를 실시하였고 각 악력 측정기 간의 차이를 알아보기 위하여 일원배치 분산분석을 실시하였으며 사후 분석은 scheffe를 이용하였다. 모든 자료는 통계 프로그램 SPSS 26.0를 이용하였으며, 유의수준은 .05 이하로 설정하였다.

## 3. 연구 결과

### 1) 연구대상자의 일반적 특성

연구 대상자는 총 49명으로 여자 28명(20.04±7.1세), 남자 21명(22.13±2.00세) 이었고 손의 길이는 여자에서 17.20±0.73cm 이었고 남자에서 19.17±0.75cm 이었다.

	Gender	N	Mean	± SD
Age(years)	F	28	20.04	0.71
	M	21	22.13	2.00
Length of hand (cm)	F	28	17.20	0.73
	M	21	19.17	0.75

Table 2. Kind of device and hand-grip strength

	Country of device	N	Hand-grip strength		F	p	Post hoc	
			Mean	± SD				
F	Right hand	28	Korea1	30.59	2.79	20.902	.000	2,3>1
			Japen2	24.67	3.81			
			China3	24.67	4.02			
	Left hand	28	Korea1	29.37	4.04	26.212	.000	2,3>1
			Japen2	22.24	3.95			
			China3	22.09	3.70			
M	Right hand	21	Korea1	35.40	6.53	9.744	.000	2,3>1
			Japen2	44.90	7.79			
			China3	46.13	7.48			
	Left hand	21	Korea1	34.63	3.46	10.004	.000	2,3>1
			Japen2	43.43	7.69			
			China3	44.03	7.31			

## 2) 분석 방법

악력 측정기의 종류별 악력 정도를 알아보았다.

여자의 오른손 악력 정도는 한국제품이  $30.59 \pm 2.79$ 로 가장 강하게 측정되었고 일본제품이  $24.67 \pm 3.81$ , 중국제품이  $24.67 \pm 4.02$  이었으며 통계학적으로 유의하게 나타났으며( $p < .001$ ), 사후검정에서는 한국제품과 일본, 중국제품이 차이가 나는 것으로 나타났다. 여자의 왼손에서 악력 정도는 한국제품이  $29.37 \pm 4.04$ 로 가장 강하게 측정되었고 일본제품이  $22.24 \pm 3.95$ , 중국제품이  $22.09 \pm 3.70$  이었고 통계학적으로 유의하게 나타났다( $p < .001$ ). 사후검정에서는 한국제품과 일본, 중국제품이 차이가 나는 것으로 나타났다.

남자의 오른손 악력 정도는 중국제품이  $46.13 \pm 7.48$ 로 가장 강하게 측정되었고 일본제품이  $44.90 \pm 7.79$ , 한국제품이  $34.63 \pm 3.46$  이었으며 통계학적으로 유의하게 나타났으며( $p < .001$ ), 사후검정에서는 한국제품과 일본, 중국제품이 차이가 나는 것으로 나타났다. 남자의 왼손에서 악력 정도는 중국제품이  $44.03 \pm 7.31$ 로 가장 강하게 측정되었고 일본제품이  $44.43 \pm 7.69$ , 한국제품이  $34.63 \pm 3.46$  순이었고 통계학적으로 유의하게 나타났다( $p < .001$ ), 사후검정에서는 한국제품과 일본, 중국제품이 차이가 나는 것으로 나타났다.

## 4. 결론

손은 일상생활에서 아주 주요한 기능을 하는 신체 도구로서 악력은 노화로 인해 매년 2%씩 감소하므로(Bassey & Harries, 1993), 신체 근력의 노화 정도와 당뇨, 인지장애와 같은 만성 질환의 예측인자로도 활용될 수 있고 산업 재해로 인한 손상 후 직업의 복귀와 직종의 선택을 위하여서도 활용되며, 아동발달 정도를 판단하는 기준으로도 사용될 수 있다.

본 연구는 두 종류의 국외 악력 측정제품과 한 종류의 국내산 악력 측정기 간의 차이를 알아보았다. 세 종류의 악력 측정기를 통해 측정한 악력 정도는 차이가 있었는데, 특히 한국제품이 일본과 중국과 차이가 나는 것으로 보인다. 한국제품에서 여자의 악력은 가장 높게 나왔고 남자는 가장 약하게 나타나서 악력의 정도에 대한 민감도가 떨어지는 것으로 보인다. 따라서, 한국 제품은 악력에 대한 민감도를 증가시키고 악력의 범위를 확장시키는 것이 필요한 것으로 보인다.

## References

- Hong S. H. et al.(2021). Relationship Between Handgrip Strength and the Prevalence of Diabetes Mellitus Among Korean Adults: Korean National Health and Nutrition Examination Survey, 2014-2018. *Korean Society of Exercise Physiology*, 30(1), 110-121. (in Korean).
- Kimi E. K. et al. (2018). Observational Evidence of the Association Between Handgrip Strength, Hand Dexterity, and Cognitive Performance in Community-Dwelling Older Adults: A Systematic Review. *Journal of Epidemiology*, 28(9), 373-381. <https://doi.org/10.2188/jea.JE20170041>
- Song, M, J. (2020). Meta-Analysis of Group Experimental Studies utilizing Psychomotor Intervention for Children with Developmental Delays. *Journal of Special Education & Rehabilitation Science*, 59(1), 27-55.
- Yoon E. S., Park S. H. (2020). Associations between Relative Handgrip Strength and Incidence of Type 2 Diabetes Mellitus in Middle-aged and Older adults: Finding from the Korean Genome and Epidemiology Study. *Korean Society of Exercise Physiology*, 29(4), 377-384. (in Korean).
- 이준석(2022). A Study on the Effects of Color Relation on Cybersickness in Virtual Reality Space, 한양대학교 박사학위논문

## 멀미 증상과 사이버 멀미와의 상관관계

노효련<sup>1</sup>, 김혜규<sup>2</sup>, 남채림<sup>3</sup>, 한재영<sup>3</sup>, 이효정<sup>3</sup>, 최종현<sup>3</sup>, 김건우<sup>3</sup>, 김유연<sup>3</sup>

### 요약

본 연구는 가상현실을 체험한 후 나타나는 사이버 멀미 증상과 멀미와의 상관관계를 알아보기 위하여 실시하였다. 연구 대상자는 20대 성인 40명으로 남녀 각각 20명이었다. 가상현실 체험 전·후에 사이버 멀미 정도를 Simulator Sickness Questionnaire를 실시하였다. 사이버 멀미의 정도에서도 남자보다 여자에서 사이버 멀미가 더 심한 것으로 나타났으며, 사이버 멀미 하부 요인 중 메스꺼움 정도는 성별에 따른 차이가 나타나지 않았으나 안구운동과 방향감각 상실 정도는 여자에서 더 심하게 나타났다. 사이버 멀미 정도와 가상현실 체험 경험 횟수와 자동차 멀미 증상 정도와는 상관관계가 나타나지 않았다. 따라서, 사이버 멀미와 가상현실 체험에 대한 경험에 대한 이론적 정립이 필요한 것으로 보인다.

주요용어 : 가상현실, 사이버 멀미, 안구운동, 방향감각 상실, 메스꺼움.

### 1. 서론

최근 COVID-19로 인한 사회적 거리두기 등의 대면 접촉의 빈도가 낮아지면서, 간접 체험을 할 수 있는 가상현실 기술에 대한 관심이 급증하였다. 가상현실 기술의 발전으로 기기가 개선되고 콘텐츠의 수의 급증이 되었으나 가상현실을 체험하며 겪게 되는 멀미감, 즉 사이버 멀미의 발생이 가상현실 체험에 큰 장애로 대두되었다. 사이버 멀미에 대한 문제를 해결하기 위하여 콘텐츠의 색상의 문제, 기기의 개선 등을 하고 있으나 사이버 멀미에 대한 문제 해결은 답보 상태이다(이준석, 2022). 여러 연구에서 사이버 멀미에 대한 해결방안을 제시하고 있으나 실제로 사이버 멀미에 대한 증상의 정도와 출현률에 대하여는 밝혀지지 않고 있다.

이에 이 연구에서는 사이버 멀미 증상 정도와 자동차 멀미와의 관계에 대하여 알아보하고자 한다.

### 2. 연구 방법

연구 대상자는 20대 성인 40명(남: 20명, 여: 20명)이었다. 가상현실 체험은 식사 후 두 시간 이후에 실시하였다. 이들은 교내에 설치되어 있는 가상현실 체험관에서 10분 길이의 가상현실 영상을 체험하도록 하였다. 사이버 멀미 정도의 측정에는 SSQ(Simulator Sickness Questionnaire)를 통해 실

<sup>1</sup>25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 교수. E-mail: withtry@kangwon.ac.kr

<sup>2</sup>(교신저자) 25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 학부재학생.

E-mail: biz1022@naver.com

<sup>3</sup>25940 강원도 삼척시 도계읍 황조길346, 강원대학교 물리치료학과 학부 재학생.

시하였으며 VR 체험 전과 후의 비교를 하였다. SSQ는 총 16문항으로 이루어져 있으며, 증상은 크게 메스꺼움(Nausea), 안구운동(Oculomotor), 방향감각 상실(Disorientation)의 세 가지 하위 척도로 나뉘고, 이 세 가지 척도는 다시 증상별로 16가지로 구분된다. SSQ의 메스꺼움, 안구운동, 방향감각 상실의 세 가지 하위 척도는 각각의 가중치에 따라 계산하여 값이 크게 나타날수록 멀미감이 증가하는 것으로 판단한다(Lee, 2022).

수집된 자료는 통계 프로그램 SPSS 21.0을 이용하였으며, 성별과 사이버 멀미는 독립표본 T-test를 차량 멀미 정도와 사이버 멀미 간에는 피어슨 상관관계를 실시하였다. 유의수준은  $p < .05$ 로 설정하였다.

### 3. 연구 결과

#### 1) 가상현실 체험 경험

가상현실 체험 경험의 횟수를 알아보았다. 처음 경험한 경우가 42.5%(17명)로 가장 많았고 10회 이상 경험한 정도는 12.5%(5명)로 나타났다. 성별에 따른 가상현실 체험 경험 횟수 차이는 나타나지 않았다.

#### 2) 성별과 사이버 멀미 정도

성별에 따른 사이버 멀미 정도를 알아보았다. 메스꺼움 정도는 성별에 따른 통계학적인 차이가 나타나지 않았으나( $p > .05$ ) 안구운동과 방향감각 상실 정도는 여자에서 더 심하게 나타났으며, 통계학적으로 유의하게 나타났다( $p < .05$ ). 전체 사이버 멀미의 정도에서도 성별의 비교에서 사이버 멀미 정도가 통계학적으로 유의하게 나타났으며( $p < .05$ ), 남자보다 여자에서 사이버 멀미가 더 심한 것으로 나타났다.

#### 3) 사이버 멀미 증상과 가상현실 체험횟수 및 멀미 정도와의 상관관계

사이버 멀미 정도와 가상현실 체험 경험 횟수와 평상시 자동차를 이용하였을 때의 평상시 멀미 증상 정도와의 상관관계를 분석하였다. 이들 변수들 간에는 상관관계가 통계학적으로 나타나지 않았다.

Table 1. Virtual reality experience

Experience (times)									n(%)	
	1	2	3	4	6	8	11	$\chi^2$	p	
M	10 50.00%	2 10.0%	2 10.0%	1 5.0%	2 10.0%	0 0.0%	3 15.0%	20 100.0%	7.396	.286
F	7 35.0%	6 30.0%	0 0.0%	0 0.0%	4 20.0%	1 5.0%	2 10.0%	20 100.0%		
Total	17 42.5%	8 20.0%	2 5.0%	1 2.5%	6 15.0%	1 2.5%	5 12.5%	40 100.0%		

Table 2. Degree of cybersickness

Symptom	Gender	Mean	SD	t	p
Nausea	M	9.06	8.46	-1.023	.313
	F	11.93	9.22		
Oculomotor	M	6.82	6.91	-2.084	.044
	F	12.89	11.03		
Disorientation	M	11.14	14.70	-2.476	.018
	F	26.45	23.42		
Total	M	9.91	7.10	-2.785	.008
	F	18.14	11.14		

Table 3. Correlation analysis with cybersickness symptoms, virtual reality experiences, and degree of sickness

	Nausea	Oculomotor	Disorientation	Total
Virtual reality experience times	0.022	-0.068	-0.023	-0.037
Car sickness	-0.077	0.05	0.138	0.073

#### 4. 결론

20대들의 가상현실에 대한 체험 경험은 50% 가량인 것으로 나타나서, 가상현실이 새로운 문화 콘텐츠로 다양한 분야에 적용되고 있으나 현재로는 확산 정도가 높은 것으로 보이지 않는다. 본 연구에서 사이버 멀미 정도는 여자들이 남자보다 높게 나타나서 성별의 차이에 대한 추가적인 연구가 필요한 것으로 보인다. 일반적으로 가상현실에 대한 체험 경험이 쌓일수록 사이버 멀미 증상이 덜 나타난다고 하였으나 본 연구에서는 상관성이 나타나지 않았으며, 자동차 멀미와도 상관성이 없는 것으로 나타났다. 이 연구에서 대상자가 40명을 많지 않았다는 제한점이 있으나, 사이버 멀미는 경험회수와 자동차 멀미와는 상관성이 없으므로 자동차 멀미에 대한 이론으로 사이버 멀미를 설명할 수 있는지에 대한 고찰과 경험에 의존하여 사이버 멀미의 감소에 대한 명확한 증거가 필요한 것으로 보인다.

#### References

Lee, G. S. (2022). A Study on the Effects of Color Relation on Cybersickness in Virtual Reality Space, Hanyang university, Dr. thesis.

Cho, H. C., Roh, H. L., Lee, H. J. (2022). The Effect of Convergence Vision Therapy on VR Cybersickness, *Journal of The Korea Convergence Society*, 13(2). 55-65.

## 원예요법을 병합한 대인간호 참여 경험

박송이<sup>1</sup>, 차지영<sup>2</sup>

주요어: 경증치매노인, 대인간호, 원예요법, 질적연구.

### 1. 연구의 필요성 및 목적

요양원 입소 경증 치매노인은 인지기능 감소와 행동심리를 경험하고 있지만 이에 적절한 간호 중재는 부족하다. 원예요법을 병합한 대인간호는 사회적 활동과 교류가 제한되어 있는 요양원 입소 경증 치매노인에게 상호작용의 대인간호와 회상이라는 긍정적 인지 작업에 초점을 둔 회상요법과 더불어 식물과의 신체적 활동이 기본이 되는 원예요법의 효과를 집단 활동을 통해 제공하는 프로그램이다. 따라서, 본 연구의 목적은 원예요법을 병합한 대인간호를 요양원 입소 경증 치매노인에게 적용한 후, 대상자의 경험을 탐색하는 것이다. 이를 통해 원예요법을 병합한 대인간호가 요양원 입소 경증치매노인의 인지기능을 증진 시키고, 행동심리증상을 감소할 수 있도록 도와주는 통합적이고, 치료적인 간호 중재로 활용될 수 있도록 기초자료를 제공하고자 한다.

### 2. 연구방법

전통적 내용분석 방법(Conventional content analysis)을 사용하여 원예요법을 병합한 대인간호에 참여한 5명의 대상자의 참여 경험을 개별 면담을 통해 탐색하였다. 본 질적 연구의 주요 질문은 “대인간호 집단회상 원예요법의 참여 후에 기분이나 감정이 어떻게 변화 되셨나요? 이다.

### 3. 연구결과

본 연구 결과에서 도출된 주제는 ‘함께 함의 중요성을 깨달음’, ‘활동참여를 통한 기쁨과 성취감을 경험함’, ‘원예가 주는 생동감과 활력을 느낌’이었다. 이러한 결과는 원예요법을 병합한 대인간호의 참여 경험이 긍정적인 경험임을 시사하고 있다.

### 4. 결론

원예요법을 병합한 대인간호를 요양원에 제공할 경우 경증 치매노인이 삶의 질을 높일 수 있도록 도와주는 통합적이고, 치료적인 간호 중재로 활용될 수 있을 것이다.

<sup>1</sup>강동대학교 간호학부 조교수.

<sup>2</sup>이화여자대학교 간호대학 부교수.

## 간호학과 신입생의 대인관계역량이 대학생활적응에 미치는 영향 2차 자료분석: 사회적 관계망 접촉빈도의 매개효과를 중심으로\*

이경은<sup>1</sup>

### 요 약

본 연구는 간호대학 신입생의 대인관계역량이 대학생활적응에 미치는 영향에서 사회적 관계망 접촉빈도의 매개효과를 확인하기 위한 서술적 조사연구이다. 2022년 5월 15일부터 31일까지 D광역시 2개 대학 간호학과 1학년에 재학 중인 141명을 대상으로 진행하였으며, SPSS macro PROCESS v.4.2\_beta의 부트스트래핑(bootstarping) 4번 모델을 활용하여 매개효과를 검증하였다. 연구 결과, 대상자의 대인관계역량은 대학생활적응에 유의한 정적 영향을 미쳤으며, 사회적 관계망 접촉빈도는 대인관계역량과 대학생활적응의 관계에서 긍정적 부분매개효과를 가졌다. 사회적 관계망 접촉빈도의 매개효과 설명력은 3.21%로 나타났으며 95% 신뢰구간에서 하한값은 .0056, 상한값은 .0665로 하한값과 상한값 사이에 0을 포함하고 있지 않아 통계적으로 유의하였다. 본 연구 결과, 간호대학 신입생의 대인관계역량이 대학생활적응에 영향을 주고 있으며, 사회적 관계망 접촉빈도가 대인관계역량과 대학생활적응 사이에서 부분매개역할을 하고 있음이 밝혀졌다. 따라서 간호대학 신입생의 대학생활적응을 향상시키기 위하여 대인관계역량과 함께 사회적 관계망 접촉빈도를 향상시킬 수 있는 대학 차원의 프로그램 개발과 지원이 중요할 것이다.

주요용어 : 대인관계역량, 대학생활적응, 사회적 관계망 접촉빈도, 매개효과

### 1. 서론

#### 1.1. 연구 필요성

대학교 신입생에게 대학생활적응이란 학교 내외에서 학업관 관련된 활동, 대인관계 형성 및 동아리 활동 등과 같은 사회생활에 주체적이고 독립적으로 적응하는 과정으로(Moon, 2018), 단순 대학생활을 즐기는 것이 아닌 취업을 위한 준비와 자아실현의 시기로 여겨진다. 특히, 코로나 19 팬데믹 상황에서 장기간 사회적 거리두기로 특징되는 고등학교 시절을 보내고 낯선 대학에 진학한 신입생들에게 대학생활의 성공적 적응은 매우 중요하다.

대학생활적응의 요소로 알려진 대인관계역량(Kim et al, 2016; Lee, Choi, 2023)은 다양한 인간관계와 사회적 상황 속에서 공동의 목표를 달성하는데 필요한 능력으로, 다른 사람의 마음, 감정 느낌을 잘 이해하고 조화롭게 살아갈 수 있는 능력을 의미한다(Gardner, Kornhaber, Wake, 1996). 사회적 관계망은 가족, 친척, 친구, 이웃 등 비공식적 사회관계망과 종교단체, 복지시설, 공공시설 등의

\*본 연구는 2022년도 계명문화대학교 교내연구비 지원으로 이루어졌음

<sup>1</sup>42601 대구광역시 달서구 달서대로 675, 계명문화대학교 간호학과 조교수. E-mail: kyungeun@kmcu.ac.kr



공식적 사회관계망을 포함하는 개념으로 대학생활적응에 영향을 미치는 요소이며(Lee, Choi, 2023), 접촉빈도가 높을수록 사회적 통합수준을 증대시키는 것으로 알려져 있다(Lee, 2013). 이는 대인관계역량의 공동의 목표를 달성하고 조화롭게 살아갈 수 있는 능력을 증진시켜 대학생활적응에 긍정적 영향을 줄 것으로 예상할 수 있다.

이에 본 연구는 간호대학 신입생에게 대인관계역량이 대학생활적응에 긍정적인 영향을 주는 과정에서, 사회적 관계망 접촉빈도의 매개효과를 파악하고자 한다.

## 1.2. 연구 목적

- 1) 간호대학생의 대인관계역량, 사회적 관계망 접촉빈도 및 대학생활적응 정도를 확인한다.
- 2) 간호대학생의 대인관계역량, 사회적 관계망 접촉빈도 및 대학생활적응 간의 상관관계를 확인한다.
- 3) 간호대학생의 대인관계역량이 대학생활적응에 영향을 미치는 과정에서 사회적 관계망 접촉빈도의 매개효과를 확인한다.

## 2. 연구 방법

### 2.1. 연구 설계

본 연구는 간호대학 신입생의 대인관계역량이 대학생활적응에 영향을 미치는 과정에서 사회적 관계망 접촉빈도의 매개효과를 확인하기 위한 서술적 조사연구이다.

### 2.2. 연구대상 및 자료수집 방법

원 자료의 대상자는 본 연구는 사전 승인받은 D 광역시 2개 대학의 간호학과 신입생을 대상으로 하였다. 본연구의 목적을 이해하고 자발적으로 참여를 동의한 대상자로 하였으며, 원 자료 수집 전 대상자의 윤리적 보호를 위하여 기관생명윤리위원회(Institutional Review Board)의 승인(DHCIRB-2022-03-0003)을 받은 후 2022년 5월 15일부터 5월 31일까지 자료 수집 후 분석을 시도하였다. 연구대상자의 일반적 특성은 다음과 같다(Table 1).

Table 1. Subjects' general characteristic (N=141)

Characteristics	Categories	n(%) or Mean±SD
Age(years)	> 19	19.65±2.86
	19 ≤	59(41.8)
	Male	82(58.2)
Gender	Female	37(26.2)
		104(73.8)

### 2.3. 연구도구

#### 1) 대인관계역량

대인관계역량 척도는 Guemey(1977)의 대인관계변화척도(RCS: Relationship Change Scale)를

Moon(1980)이 번역한 것을 Lee(2014)가 수정한 문항을 사용하였다. 본 척도는 총 18문항, 5점 척도로 점수가 높을수록 대인관계역량이 높음을 의미한다. Lee(2014)의 연구에서 신뢰도는 Cronbach's  $\alpha = .80$ 이었으며, 본 연구에서는 Cronbach's  $\alpha = .90$ 이었다.

## 2) 사회적 관계망 접촉빈도

사회적 관계망은 Tracy와 Whittaker(1990)가 개발한 사회적 관계망 지도(Social Network amp)를 Kim, Choi(2019)의 연구에서 번안 및 수정한 것을 토대로 하였다. 사회적 관계망은 동거가족, 가족 및 친척, 친구, 직장 및 학교, 종교 및 클럽, 이웃, 전문 서비스 제공자의 7개 영역으로 구분하여 접촉빈도를 측정하였다. 여기서 접촉빈도는 대면접촉, 전화, 편지 등을 이용하여 사람들과 접촉하는 정도를 의미한다. 각 유형별 접촉 횟수의 근사치는 Lee(2020)의 연구를 근거로 하여 사회적 관계망 접촉빈도를 산출하였다. 거의 만나지 않음 0, 일년에 1번 0.1, 일년에 2~3번 0.2, 한달에 한번 1, 한달에 2~3번 2.5, 일주일에 1번 4, 일주일에 2~3번 10, 거의 매일 20을 부여하고 이를 합산한 값이다.

## 3) 대학생활적응

대학생활적응은 Baker, Siryk(1984)이 개발한 대학생활적응도검사(SACQ)를 Lee(1999)가 수정·보완한 검사지를 사용하여 측정하였다. 본 척도는 총 25문항, 5점 척도로 점수가 높을수록 대학생활적응을 잘 하고 있음을 의미한다. Lee(1999)의 연구에서 신뢰도는 Cronbach's  $\alpha = .85$ 였으며, 본 연구에서는 Cronbach's  $\alpha = .84$ 였다.

## 2.4. 자료분석방법

대상자의 일반적 특성, 대인관계역량, 사회적 관계망 접촉빈도 및 대학생활적응 정도는 SPSS Win 21.0 program을 이용하여 빈도 분석, 기술 통계분석을 하였으며, 매개효과는 SPSS PROCESS v.4.2\_beta의 부트스트래핑(bootstarpping) 4번 모델을 활용하여 검증하였다.

# 3. 연구 결과

## 3.1. 대상자의 대인관계역량, 사회적 관계망 접촉빈도 및 대학생활적응 정도

대상자의 대인관계역량은 평균 3.96점, 사회적 관계망 접촉빈도는 평균 51.10점, 대학생활적응은 평균 3.60점으로 나타났다(Table 2).

Table 2. Subjects' general characteristisc (N=141)

Variable	Range	Min	Max	Mean±SD
Interpersonal Relation Competency	1~5	2.06	4.98	3.96±0.53
Social Network Contact Frequency	0~160	22.60	81.0	51.10±11.60
College Life Adaptation	1~5	2.28	4.84	3.60±0.46

### 3.2. 대인관계역량, 사회적 관계망 접촉빈도 및 대학생활적응 간의 상관관계

대상자의 대인관계역량은 사회적 관계망 접촉빈도( $r=.24, p<.001$ ) 및 대학생활적응( $r=.71, p<.001$ )과 유의한 양의 상관관계를, 사회적 관계망 접촉빈도는 대학생활적응( $r=.31, p<.001$ )과 유의한 양의 상관관계를 나타내었다(Table 3).

Table 3. Correlation between interpersonal relation competency, social network contact frequency, and college life adaptation(N=141)

Variable	Interpersonal Relation Competency	Social Network Contact Frequency	College Life Adaptation
Interpersonal Relation Competency	1		
Social Network Contact Frequency	.24(<.001)	1	
College Life Adaptation	.71(<.001)	.31(<.001)	1

### 3.4. 대인관계역량이 대학생활적응에 영향미치는 과정에서 사회적 관계망 접촉빈도의 매개효과

대상자의 대인관계역량과 사회적 관계망 접촉빈도를 동시투입하였을 때 대인관계역량의 영향력이 유의하게 나타났으며( $B=.67, p<.001$ ), 사회적 관계망 접촉빈도도 유의하게 나타났다( $B=.16, p=.011$ ). 사회적 관계망 접촉빈도의 부분매개효과가 확인되었으며 대인관계역량은 대학생활적응에 직접적인 영향을 미치며, 사회적 관계망 접촉빈도를 통해서도 간접적으로 영향을 미치는 것으로 나타났다(Table 4).

매개효과의 통계적 유의성 검증 결과, 매개효과 계수는 .032이며, 95% 신뢰구간에서 하한값(LLCI)과 상한값(ULCI) 사이에 0이 포함되지 않아 신뢰도 .05수준에서 유의하다(Table 5).

Table 4. The mediating effect of social network contact frequency between interpersonal relation competency and college life adaptation(N=141)

Independent variable	Dependent variable : Social Network Contact Frequency				Dependent variable : College Life Adaptation			
	B	S.E	t	p	B	S.E	t	p
Constant	30.613	7.188	4.259	<.001	.961	.218	4.412	<.001
Interpersonal Relation Competency	.237	1.797	2.876	.0047	.672	.053	11.130	<.001
Social Network Contact Frequency					.155	.002	2.569	.011
	$R^2 = .056, F=8.268, p=.005$				$R^2 = .525, F=76.30, p<.001$			

Table 5. Bootstarping results of social network contact frequency(N=141)

Variable	Indirect Effect	Boot SE	95% 신뢰구간	
			LLCI	ULCI
Social Network Contact Frequency	.032	.016	.006	.067

## 4. 결론 및 제언

본 연구 결과, 간호대학 신입생의 대인관계역량이 대학생활적응에 영향을 주고 있으며, 사회적 관계망 접촉빈도가 대인관계역량과 대학생활적응 사이에서 부분매개역할을 하고 있음이 밝혀졌다.

따라서 간호대학 신입생의 대학생활적응을 향상시키기 위하여 대인관계역량과 함께 사회적 관계망 접촉빈도를 향상시킬 수 있는 대학 차원의 프로그램 개발과 지원이 중요할 것이다.

## References

- Baker, Siryk (1984). Measuring adjustment to college. *Journal of Counseling Psychology*, 31(2), 179-189.
- Gardner, H., Kornhaber, M. L., & Wake, W. K. (1996). *Intelligence: Multiple perspectives*. Fort Worth: Harcourt Brace College Publishers.
- Guernsey Jr, B. G. (1977). *Relationship enhancement: Skill training programs for therapy, problem prevention, and enrichment*. San Francisco: Jossey-Bass.
- Kim, M. A., Lee, K. H., Park, H. J., Kim, E. J. LEE I. H. (2016). Relationships among Irrational Beliefs, Self-Efficacy, and Adjustment to College with Four-Year-Course Nursing Freshmen: Mediating Effect of Self-Efficacy. *Journal of The Korean Data Analysis Society*, 18(2), 1163-1178.
- Lee, J. E. (2014). *The Structural Modeling on Interpersonal Competence of Nurses*. Doctor's thesis, Keimyung University. (in Korean).
- Lee, K. E., Choi, H., S. (2023). Effect of Social Network, Interpersonal Relation Competency and Problem Solving Competency of Nursing College Freshmen on College Life Adaptation in the Post-Corona Era, *Journal of Learner-Centered Curriculun and Instruction*. 23(2), 103-113.
- Lee, M. A. (2013). The U-shaped Relationship Between the Size of Social Networks and Depressive Symptoms: Examining the Moderating Effect of Interpersonal Trust. *Korean Journal of Sociology*, 47(4), 171-200.
- Lee, S. W. (2020). *The Relationship between Internalized Shame and Loneliness: The Mediating Effect of Self-Concealment and Perceived Social Network*. Master's thesis, Busan University. (in Korean).
- Lee, Y. J. (1999). *The Non-Residential Student's Adaptation to College Life and Career Plans*. Master's thesis, Ewha Womans University. (in Korean).
- Moon, H. J. (2018). The Development and Validation of the College Adjustment Scale for Freshmen majoring in Airline Service, *Journal of Tourism and Leisure Research*. 30(11), 75-94.
- Tracy, E. M., & Whittaker, J. K. (1990). The social network map: Assessing social support in clinical practice. *Families in Society*, 71(8), 461-470.

## Importance of Multimorbidity in Predicting Suicidal Ideation Among the Older Population Using Four Machine Learning Methods

*Young-Bin Seo*<sup>1</sup>, *Hae-Young Kim*<sup>2</sup>, *Ki-Bong Cho*<sup>3</sup>

### Abstract

This study aims to investigate the importance of specific combinations of chronic disease, multimorbidity, in predicting suicidal ideation (SI) among the older Korean population using machine learning (ML) technique. The analysis utilized data from the Korea National Health and Nutrition Examination Survey (KNHANES) conducted between 2010 and 2019, involving 9,008 individuals aged over 65 years. Recursive feature elimination with cross-validation (RFECV) was employed to select the most relevant variables. The average age of the participants is 72.7 years(SD:5.0), 45.5% being men. Suicidal ideation was reported by 12.4% of the participants, among whom 3.5% had one chronic disease and 6.8% had multimorbidity. The performance of the four ML models, as measured by the receiver operating curve area under the curve (AUC), ranged from 0.714 (random forest)-0.831 (logistic regression, LR). LR model exhibited a sensitivity of 0.287 and precision of 0.582. Among the individual chronic disease with top 30 feature importances, the following conditions (rank, OR) showed an increased odds of SI : stroke (13, 1.16), myocardial infarction (14, 1.67), osteoarthritis (16, 1.20), asthma (18, 1.64), diabetes (19, 1.12), hypertension (21, 1.11), and angina (22, 1.4) by the order of importance. Notably, specific combinations of multimorbidity, including two and three combination among osteoarthritis, diabetes, hypertension, angina, cancer, asthma and dyslipidemia, were selected in top 30 importance features associated with an increased risk of SI. The finding suggest that the LR model can serve as an effective decision-making tool for accurately predicting suicide and identifying early suicide risk in older polulations with a high prevalence of multimorbidity.

*Keywords* : Suicidal Ideation, Multimorbidity, Machine Learning, Feature Importance.

---

<sup>1</sup>Graduate student, Department of Public Health Sciences, Graduate School, and BK21 Four R&E Center for Learning Health Systems, Korea University, 45 Anam-ro, Seongbuk-gu, Seoul 02841, Korea.  
E-mail: tjdudqls98@gmail.com

<sup>2</sup>(Corresponding Author) Professor, Department of Public Health Sciences, Graduate School, and BK21 Four R&E Center for Learning Health Systems, Korea University, 45 Anam-ro, Seongbuk-gu, Seoul 02841, Korea. E-mail: kimhaey@korea.ac.kr

<sup>3</sup>Visiting Professor, Department of Health Policy and Management, College of Health Science, Korea University, 45 Anam-ro, Seongbuk-gu, Seoul 02841, Korea. E-mail: kbchoi00@korea.ac.kr

## 간호대학생의 감성지능, 의사소통능력, 대인관계유능성 및 대학생활적응의 관계

이은미<sup>1</sup>, 오윤정<sup>2</sup>

### 요 약

본 연구는 간호대학생의 감성지능, 대인관계유능성 및 대학생활적응 간의 관계를 파악하고 대학생활적응에 영향을 미치는 요인을 확인하기 위한 서술적 조사연구이다. 연구대상자는 G시 소재 대학에 재학 중인 간호대학생 173명을 대상으로 하였으며, 자료수집은 자가보고식 설문지를 이용하여 2022년 12월 3일에서 12월 20일까지 실시하였다. 수집된 자료는 SPSS 25.0 program을 이용하여 기술적 통계, t-test, ANOVA, Pearson correlation coefficient, 다중회귀분석을 시행하였다. 연구결과 대학생활적응은 전공만족도, 간호학과 지원동기에 따라 유의한 차이를 보였으며, 감성지능( $r=.760, p<.001$ ), 의사소통능력( $r=.600, p<.001$ ), 대인관계유능성( $r=.451, p<.001$ ) 간에 유의한 양의 상관관계를 나타냈다. 대학생활적응에 유의한 영향을 미치는 요인은 감성지능( $\beta=.543, p<.001$ ), 의사소통능력( $\beta=.433, p=.001$ ), 대인관계유능성( $\beta=.283, p<.031$ )으로 나타났다. 이 변인들의 대학생활적응에 대한 총 설명력은 59.9%였으며 감성지능이 가장 주된 영향요인임이 확인되었다. 따라서 간호대학생의 감성지능과 의사소통능력을 함양하고 대인관계유능성 향상을 돕는 교육프로그램을 개발하고 간호학과 교육과정 동안 지속적으로 제공함으로써 대학생활적응력을 높일 수 있는 전략이 필요하다.

주요용어 : 감성지능, 의사소통능력, 대인관계유능성, 대학생활적응, 간호대학생.

<sup>1</sup>광주광역시 광산구 침단중앙로 23, 남부대학교 간호학과 조교수. E-mail: jjang9312@nambu.ac.kr

<sup>2</sup>광주광역시 광산구 침단중앙로 23, 남부대학교 간호학과 조교수. E-mail: Agnes@nambu.ac.kr

## 간호대학생의 인권의식, 공감능력 및 다문화수용성과의 관계

오윤정<sup>1</sup>, 이은미<sup>2</sup>

### 요 약

본 연구는 간호대학생의 인권의식, 공감능력 및 다문화수용성 간의 관계를 파악하기 위한 서술적 조사연구이다. 연구대상자는 G시 소재 대학에 재학 중인 간호대학생 120명을 대상으로 하였으며, 자료수집은 자가보고식 설문지를 이용하여 2022년 10월 10일에서 11월 5일까지 실시하였다. 수집된 자료는 SPSS 22.0 program을 이용하여 기술적 통계, t-test, ANOVA, Pearson correlation coefficient을 시행하였다. 연구결과 대상자의 평균 나이는  $21.08 \pm 3.30$ 세였으며, 해외여행경험은 70명(58.3%)가 있었고, 해외거주 경험은 7명(5.8%)가 있었다. 외국인이나 이주민과의 교류경험은 56명(46.7%)가 있었고, 외국어 구사 능력은 43명(35.8%)가 중간정도였으며 간호윤리교육경험은 85명(70.8%)가 있었다. 다문화인식은 간호윤리경험( $p=.003$ )에 따라 유의한 차이를 보였으며, 인권의식과 공감능력( $r=.249, p=.006$ )은 서로 유의한 양의 상관관계가 있었으나, 다문화수용성과는 ( $r=.030, p=.744$ ) 서로 유의한 상관관계를 보이지 않았다. 본 연구는 간호대학생의 인권의식, 공감능력 및 다문화수용성 간 관계를 파악함으로써 간호대학생의 인권의식과 공감능력 및 다문화수용성 간 상관성을 파악했다는 점에 의의가 있다. 따라서 간호대학생의 다문화수용성에 영향을 미치는 광범위한 요인을 탐색하기 위한 연구를 제안하며 간호대학생의 인권의식, 공감능력 및 다문화수용성을 증진시킬 수 있는 효과적인 교육적 중재의 개발과 적용을 제안한다.

주요용어 : 인권의식, 공감능력, 다문화수용성, 간호대학생, 다문화.

<sup>1</sup>광주광역시 광산구 첨단중앙로 23, 남부대학교 간호학과 조교수. E-mail: Agnes@nambu.ac.kr

<sup>2</sup>광주광역시 광산구 첨단중앙로 23, 남부대학교 간호학과 조교수. E-mail: jjang9312@nambu.ac.kr





# 트렌드리서치(Trendresearch)



## 트렌드리서치는?

국내&해외 마케팅 및 사회/여론 조사 경력이 20년 넘는 전문가가 직접 연구에 참여해 통찰력 있는 INSIGHT 제공  
트렌드리서치의 약 72만 ACCESS PANEL을 통한 정확한 조사 수행이 가능함



마케팅 전문가와  
사회조사 전문가의 결합



자사 패널을 직접 보유  
(720,000명)



별도의 정성조사 전문가 활용

## 주요 서비스



온라인/모바일 조사

전문인력을 통해 국내 온라인 및  
모바일 조사를 진행하고 있으며  
Quick survey, Mobile Diary  
등의 조사 수행



마케팅/소비자 조사

마케팅 의사 결정을 위한 시장  
및 소비자 정보 제공



학술/연구/석박사 논문 조사

학술/연구/석박사 논문의 경우  
사회기여 차원으로 일반 조사와는  
차별화된 단가(다운된 단가)로 제공

## T-Panel 현황

### 해외조사

해외 제휴 네트워크를 활용하여  
신속한 조사 진행이 가능합니다.

### 특정 포탈 사이트 회원이 아닙니다

조사참여 횟수, 조사 참여일,  
참여조사 종류 및 응답의  
성실도를 체계적으로  
관리합니다.

### 대표성이 있습니다

샘플링을 통해 추출된 표본들에게만  
이메일, 문자, 알림톡을 발송하여  
조사를 진행합니다.

### 분석 서비스

조사 결과의 통계 검정  
결과를 제공하여 분석의  
신뢰성을 제고합니다.

Available  
Panels  
720,000

SAS Korea

 sas viya

## A faster, more productive AI and analytics platform

What if you could make decisions more decisively?  
And your team could work more productively?

You can with SAS® Viya. SAS is committed to creating technology that is not only collaborative, powerful and intuitive, but also ethical, equitable and sustainable. Together, we can build a better, more productive future for all.

[sas.com/viya](https://sas.com/viya)



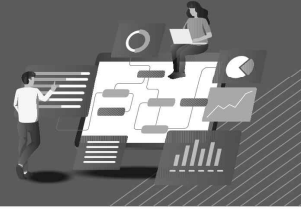
SAS® OnDemand for Academics

- SAS 클라우드 기반 소프트웨어를 무료로 체험해보십시오.
- 먼저 SAS Profile을 생성하고 Sign Up 해주십시오.

# 데이터솔루션

## 데이터솔루션 | 前 SPSS Korea

데이터솔루션은 Digital Transformation Journey를 함께 하는 Cloud와 Data 기반의 Digital Innovation Provider입니다. 데이터솔루션은 Data와 Application Lifecycle 전 과정에 최적화된 Total Service와 솔루션을 보유하고 있으며, 특히 Cloud, IT Modernization, Analytics, Intelligence Service를 통해 고객에게 새로운 Insight와 차별화된 Value를 제공하고, 고객의 비즈니스 성공을 위한 Digital Innovation을 함께 합니다.



### 주요 분석 사업 영역

SPSS Korea를 전신으로 한 Data & Analytics 부문에서는 통계소프트웨어/ML/AI 솔루션 공급 및 데이터분석 컨설팅 사업을 수행하며, 고객의 데이터 기반 의사결정을 지원하고 있습니다. 30여 년간 제조, 금융, 유통, 통신 등 다양한 산업 군에서 경험 및 지속적인 연구개발을 통해 통계분석에서 ML/AI로 진화하는 시장의 변화를 이끌고 있습니다.

Analytics Consulting	MAaaS (Managed Analytics as a Service)	Intelligence Solution	AI/BigData Education
<ul style="list-style-type: none"> <li>· 수요 예측</li> <li>· 리스크 예측</li> <li>· 상품 추천</li> <li>· 품질 분석 및 예지 보전</li> </ul>	<ul style="list-style-type: none"> <li>· MAaaS PASP</li> <li>· PoC</li> <li>· Pro</li> <li>· Maintenance</li> <li>· MAaaS by AI Model</li> </ul>	<ul style="list-style-type: none"> <li>· KoreaPlus Statistics (Embaded on SPSS)</li> <li>· KNIME</li> <li>· IBM CP4D</li> <li>· Market Mind</li> <li>· Brightics AI</li> <li>· AI.NER</li> </ul>	<ul style="list-style-type: none"> <li>· 빅데이터링센터</li> <li>· SPSS/Python/R 교육 연간 80여개</li> <li>· 온/오프라인 강의제공</li> </ul>

### 데이터솔루션 분석 사례

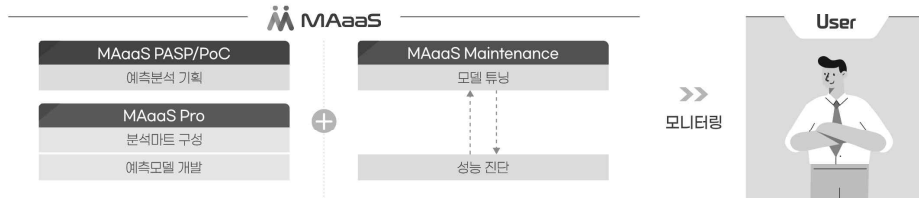
<b>수요 예측</b> 제조   수요예측 분석 시스템 구축 유통   판매량 예측 모델 개발 물류   물동량 예측 시스템 구축 공공   열 수요 예측 모델 개발	<b>텍스트 분석</b> 통신   VoC 분석 금융   텍스트 분석을 활용한 평판 분석 유통   텍스트 분석을 통한 마케팅 증권   챗봇 메시지 분석
<b>상품 추천</b> 유통   AI 상품 추천 프로그램 구축 유통   온라인 쇼핑물 개인화 상품 추천 모델 개발 유통   개인화 추천 서비스 모델 구축 금융   고객관리 시스템을 통한 상품 추천	<b>품질 분석 및 예지보전</b> 제조   취약 부위 예측 모델 개발 제조   특정 공정이상 예측 분석 제조   공정 최적화 및 설비 고장 모델 개발 화학   수율 예측 시뮬레이터 개발
<b>이미지 분석</b> 제조   이미지 데이터 기반의 양불 판정 제조   비전 검사를 통한 양불 판정	

### AI 예측서비스, 이제 구독하세요!

No 플랫폼   No 데이터수집   No 데이터/사이언티스트   No 유지관리

#### MAaaS | Managed Analytics as a Service

MAaaS는 AI 예측모델 기획부터 유지관리까지의 Data Life-cycle에서 필요한 컨설팅을 제공하는 데이터솔루션의 AI 예측분석 토털 서비스입니다. 이제는 분석 예측 결과를 구독으로 제공받으세요.



# 지디에스컨설팅그룹

## Company : 회사소개 및 연혁

2003년 이래, 데이터 가공/분석 및 유통의 신시장 개척에 주력하였습니다.

회사명 ㈜지디에스컨설팅그룹

설립일 2015년 2월

대표이사 김은석

업종 데이터베이스, 소프트웨어 자문, 개발 및 공급, 전자상거래업

주소 서울 성동구 광나루로 228, 9층 (성수동2가, 렉스모터스)

연락처 Tel: 02-2135-8895(대표)  
Fax: 02-2135-8896



2018.07  
고려대 '데이터 사이언스 센터' 업무 제휴 (교육 및 인턴십)

2017.07  
KB국민카드 '빅데이터 센터' 업무 제휴 (카드 데이터 상품개발)

2016.09  
KT '빅데이터 지원단' 업무 제휴 (빅데이터 가공, 중개, 서비스 분야)

2016.08  
데이터 거래 중개 사업 개시 (미래창조과학부 ICT 기금사업)  
부실연구소 '한국공공정보개발원' 설립 (산기협 2016113799호)

2016.07  
빅데이터 분석 사업 개시 (지속가능한 빅데이터 사업, 경기도)

2015.04  
국가 DB 구축 사업 개시 (유통연구 실사 DB 구축, NIA)

2015.02  
㈜지디에스컨설팅그룹 설립  
(㈜지디에스케이 내 컨설팅 사업본부의 법인 분할)

2004.10  
벤처기업 등록 (041134221-1-01033호)

2003.02  
㈜지디에스케이 설립 (대표이사 김은석 외)

## Business : 사업 영역

데이터 활용 및 유통의 일관된 서비스 제공!!

사업영역	빅데이터 분석 컨설팅	데이터 가공	빅데이터 플랫폼 구축	데이터 판매/유통
사업영역		자세히 보기 >	자세히 보기 >	자세히 보기 >
상품 및 서비스	<ul style="list-style-type: none"> <li>데이터 관련 ISP 및 계획수립</li> <li>주요 빅데이터 (통신, 카드) 공급 및 활용</li> <li>공공 데이터 활용 서비스</li> <li>데이터 분석 및 시각화 서비스</li> </ul>	<ul style="list-style-type: none"> <li>공공 개방 정보의 정제, 가공 및 제공</li> <li>통신, 카드 등 빅데이터 정제, 결합 및 제공</li> <li>비정형 데이터 가공 및 제공</li> </ul>	<ul style="list-style-type: none"> <li>통계분석 도구 (알트릭스) 판매</li> <li>데이터 융복합 및 통합 시스템 구축</li> <li>빅데이터 분석 시스템 구축</li> </ul>	<ul style="list-style-type: none"> <li>데이터 사용영역별 데이터 판매</li> <li>업무 영역별 맞춤형 데이터 구축 및 제공</li> <li>주제별 맞춤형 데이터 제공</li> </ul>
사업사례	<ul style="list-style-type: none"> <li>국방기술품질원, 빅데이터 시범사업 (17-)</li> <li>경기도, 지속가능한 빅데이터 분석 (16)</li> <li>서울시, GRDP 분석 모델링 구축 (16)</li> <li>건강보험공단, 빅데이터 노인질환 모델 구축 (14)</li> <li>서울시, 유통연구 DB 구축 및 모델링 (15)</li> </ul>	<ul style="list-style-type: none"> <li>삼성전자, 법안영입 데이터 제공 (15-)</li> <li>한성, 가구판매 고객 데이터 제공 (15-)</li> <li>코웨이, 고객정보 관리 데이터 제공 (06-)</li> <li>KT, 지사 관리 데이터 제공 (10-)</li> <li>기업은행, 영업점 관리 데이터 제공 (10-)</li> </ul>	<ul style="list-style-type: none"> <li>청원시, 빅데이터 분석시스템 구축 (17)</li> <li>KT, AI 조류독감 시각화 시스템 구축 (16)</li> <li>경상북도, 통계정보시스템 구축 (15)</li> <li>KT, 지사 관리 데이터 제공 (10-현재)</li> <li>서울시, 도시통계 시스템 구축 (14, 15)</li> </ul>	<ul style="list-style-type: none"> <li>한국정보화진흥원, 데이터 거래 중개 플랫폼 구축 (16)</li> <li>한국정보화진흥원, 소상공인 데이터 제공 플랫폼 (17)</li> <li>경상북도, 서비스 인구구축을 위한 데이터 (16)</li> <li>코웨이, 고객정보 관리를 위한 데이터 (16) 외 다수</li> </ul>