



ISSN: 3022-5388

JKAI website: <https://accesson.kr/jkaia>DOI: <http://dx.doi.org/10.24225/jkaia.2024.2.2.39>

1인 가구와 다인 가구의 연령대별 위장질환 발생률 예측 모델

Predictive Model for Age-specific Incidence Rates of Gastrointestinal Diseases in Single-Person and Multi-Person Households

Na-Yeon KANG¹, Kyung-A KIM²

Received: November 05, 2024. Revised: December 06, 2024. Accepted: December 12, 2024.

Abstract

This study employed machine learning and deep learning regression models to predict the incidence of gastrointestinal diseases. In particular, attention was given to the increasing number of single-person households, emphasizing that irregular eating habits and dining out may contribute to gastrointestinal health problems. The incidence of gastrointestinal diseases was predicted and compared between single-person and multi-person households by age group, with categorical variables such as age, gender, disease, and household type converted into numerical data through one-hot encoding during the data preprocessing phase. The regression models used for prediction included linear regression, random forest, support vector machines (SVM), gradient boosting, XGBoost, and multilayer perceptron (MLP). After constructing the models, their performance was evaluated to identify the most suitable one. The evaluation results indicated that the XGBoost regression model provided the best performance. Using this model, predictions of gastrointestinal diseases by age group for both single-person and multi-person households were made and visualized. The findings highlight the need for dietary improvement programs and policy development for single-person households and are expected to contribute to the formulation of future public health policies.

Keywords : Single person household, Multi-person household, Gastrointestinal disease, Machine learning, Regression

Major Classification Code : Artificial Intelligence, Machine Learning, Prediction Analysis

1. Introduction

최근 우리나라는 1인 가구의 증가 현상이 두드러지고 있다. 통계청에 따르면, 1인 가구의 비율이 2010년 24.4%에서 2023년 약 35.5% 수준으로 확대되었고, 향후 2052년에는 약 41.3%로 높아질 것으로 전망한다.

1인 가구가 증가함에 따라 식습관이 변화하고 있다. 1인 가구는 대부분 혼자 밥 먹는(혼밥) 비중이 높다. 연구에 따르면 청년 1인 가구의 43.7%가 하루 한 번 이상 혼자

식사하고, 17.1%가 두 번 이상 혼자 식사하는 것으로 보였다. 이들은 잦은 혼밥으로 결식이나 편식과 같은 불규칙한 식습관이 자리 잡기 쉽다. 1인 가구에 관한 한 연구에서는 1주일 중 식사를 거른 횟수가 평균 2.39일이며, 1년간 식사를 거른 적이 있는 날이 11.38%에 달하는 것으로 보고했다.

또한, 혼자 밥을 먹으면 함께 밥 먹는 상대가 없기 때문에 주로 티비를 보거나 인터넷, 게임을 하면서 식사한다. 이런 식습관은 식사 속도를 빠르게 하고 음식물을 제대로 씹지

1 First Author. Undergraduate student, Big data medical convergence, Eulji University, South Korea. Email: nayeon5948@g.eulji.ac.kr

2 Corresponding Author. Professor, Big data medical convergence, Eulji University, South Korea, Email: kyungakim@naver.com

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

않아 식욕 억제 호르몬의 분비가 어려워 과식을 초래하기 쉽다. 결식과 과식 같은 불규칙한 식습관은 소화 불량과 위염, 위궤양 같은 위장질환 발생 위험을 높이는 요인으로 작용한다.

1인 가구의 급증은 식습관 변화뿐만 아니라 식품 및 외식 산업계에도 변화를 불러왔다. 즉석요리, HMR 식품 등과 같은 솔로 이코노미 형태의 상품이 증가하고, 배달 및 테이크아웃 시장이 확대되었다. 1인 가구는 조리 과정이 불필요하다는 이유로 편의점 식품이나 외식 및 배달, 테이크아웃 음식을 이용하는 경우가 많다. 2023년 식품소비행태 보고서에 따르면 주 1일 배달/테이크아웃을 이용하는 1인 가구가 41.3%로 다른 가구에 비해 10%p 이상의 차이를 보였다. 또한, 배달/테이크아웃을 이용하는 빈도가 1인 가구에서 가장 높았으며, 이러한 형태의 식생활에 의존하고 있다는 결과를 보였다. 1인 가구가 즐겨 먹는 외부 음식들은 집밥에 비해 맵고, 짜고, 자극적인 특징을 갖고 있다. 이는 지속적으로 위 점막을 자극해 점막이 얇아지는 위축성 위염을 유발할 수 있다.

최근 1인 가구의 식습관 문제점들이 대두되면서 이와 관련된 연구가 점차 늘어나고 있다. 대부분의 연구는 1인 가구와 다인 가구의 여러 요인에 따른 식생활 관계를 분석하는 데 초점을 맞추고 있다. 그러나 연령대별로 다인 가구와 1인 가구의 위장질환 발생률에 대한 구체적인 정보는 제한적이다.

따라서 본 연구는 1인 가구와 다인 가구의 위장질환 발생률을 예측하고, 이를 비교 분석함으로써 불규칙한 식습관을 가진 1인 가구에게 위장질환의 위험성에 대한 경각심을 불러일으키고, 건강한 식습관의 중요성을 강조하고자 한다.

2. Body

2.1 Research Objective

이 연구의 목적은 연령, 성별, 연도, 다인 가구 수, 1인 가구 수, 위장질환 진단 데이터에 기반하여, 1인 가구와 다인 가구 각각의 위장질환 예측 모델을 구축하고 이를 비교하는 것이다. 또한, 1인 가구와 다인 가구의 연령별 위장질환 발생률과 질환 종류를 비교 분석한다. 이 연구를 통해 연령대별 위장질환 발생률과 질환 유형을 파악하고, 이를 바탕으로 1인 가구의 건강한 식습관을 제안하고자 한다.

2.2. Research Data

본 연구에서는 위장질환 데이터를 KOSIS(통계청)에서 제공하는 연령별, 성별 통계 자료 중 위암, 대장암, 간암, 폐암 관련 문항을 활용하였다. 구체적으로, 위장질환 진단 여부와 관련된 문항만을 선택하여 각 질환의 발생률에 대한 정보를 수집하였다. 이 데이터는 연령대에 따른 위장질환 발생 경향을 파악하는 데 중요한 기초 자료로 사용된다.

1인 가구와 다인 가구 데이터는 KOSIS의 가구주 성별, 연령 및 세대 구성별 통계를 사용하였으며, 다인 가구는 일반 가구 문항을, 1인 가구는 1인 가구 문항을 선택하여 비교 분석의 기반을 마련하였다.

공동 변수로는 연령, 성별, 연도를 설정하였고, 연도는 2018년부터 2022년, 성별은 남성, 여성으로 구분하였다. 연령은 "20세~24세"부터 "85세 이상"까지 5세 간격으로 나눈 총 14개의 구간으로 설정하였다.

위장질환 변수로는 위궤양, 위축성 위염, 장상피화생, 위용종의 4가지 질환을 구분하여 사용하였다. 위궤양은 위점막이 헐어서 근육층까지 침범한 상태이며, 위축성 위염은 위점막이 위축되어 위산 분비가 감소하는 만성 위염의 일종이다. 장상피화생은 만성 위염으로 인해 위점막이 장점막처럼 변한 상태를 말하며, 위용종은 위점막의 상피세포가 자라 표면 위로 솟아오른 버섯 모양의 혹이다.

2.3. Variable Setup

모델 학습에 사용할 독립 변수(X)와 종속 변수(y)를 설정하였다. 독립 변수는 1인 가구 수, 다인 가구 수, 연령, 성별, 질병, 가구 형태로 구성되었으며, 종속 변수는 질병 발생률로 설정하였다. 범주형 변수인 독립 변수는 원-핫 인코딩을 통해 수치형 데이터로 변환되었다. 이 과정은 모델이 범주형 변수를 효과적으로 처리할 수 있도록 돕는다.

3. Research Methods and Analysis Results

3.1. Models

본 연구에서는 1인 가구와 다인 가구의 연령대별 질병 발생률을 예측하기 위해 회귀 기반 알고리즘을 사용하였다. 회귀 모델은 연구의 목표인 질병 발생률과 같은 연속적이고 실수 형태의 값을 예측하기에 가장 적합한 방식이다. 또한, 다양한 인구학적 특성(예: 가구 형태, 연령대 등)과 질병 발생률 간의 관계를 학습하여, 독립 변수와 종속 변수 간의

상관성을 효과적으로 반영할 수 있다는 점에서 최적의 방법이다.

다양한 딥러닝 및 머신러닝의 회귀 모델을 구축하고, 주어진 데이터와 연구 목표에 적합한 모델을 선정한다. 이를 위해 학습에 사용될 모델은 선형 회귀 모델, 랜덤 포레스트 회귀 모델, 서포트 벡터 회귀 모델, 그래디언트 부스팅 회귀 모델, XGBoost 회귀 모델, 다층 퍼셉트론 회귀 모델로 총 6 가지의 모델을 구축하고, 성능 평가를 진행한다.

3.1.1. Linear Regression Model

종속 변수(y)와 하나 이상의 독립 변수(X) 간의 선형 관계를 가정하여, 이를 통해 종속 변수의 값을 예측하는 모델이다. 선형 회귀의 기본 가정은, 독립 변수의 변화가 종속 변수의 선형적인 변화로 이어진다는 것이다.

3.1.2. Random Forest Regression Model

여러 개의 이진 트리를 결합하여 예측을 수행하는 앙상블 학습 방법 중 하나이다. 각각의 트리는 훈련 데이터에서 무작위로 샘플링하여 학습되며, 최종 예측은 모든 트리의 예측 결과를 평균하여 산출한다. 과적합을 방지하고, 복잡한 데이터의 비선형적 관계를 효과적으로 학습할 수 있어, 다양한 데이터에 강력한 예측 성능을 보인다.

3.1.3. Support Vector Regression

주어진 입력 데이터에 대해 목표 값과의 오차가 허용된 범위 내에 있도록 데이터를 예측하는 모델로, 초평면을 통해 연속적인 값을 예측하는 회귀 알고리즘이다. 예측 시 마진을 최대화하여 모델의 일반화 성능을 높이는 것이 특징이다.

3.1.4. Gradient Boosting Regression Model

여러 개의 약한 학습기를 순차적으로 결합하여 높은 예측 성능을 달성하는 앙상블 학습 기법 중 하나이다. 이전 모델이 학습한 결과의 오차를 줄이는 방식으로 학습되며, 최종 예측 값은 모든 모델의 가중합으로 계산된다. 복잡한 데이터에서의 높은 예측 정확도를 제공하고, 하이퍼파라미터 조절을 통해 성능을 최적화할 수 있다.

3.1.5. XGBoost Regression Model

그래디언트 부스팅을 기반으로 한 앙상블 학습 기법으로, 높은 예측 성능과 빠른 학습 속도를 제공한다. 정규화를 통해 과적합을 방지하고, 대규모 데이터에서도 효율적으로 학습할 수 있다.

3.1.6. Multilayer Perceptron Regression Model

인공 신경망을 기반으로 한 회귀 모델로, 여러 개의 은닉층을 통해 복잡한 비선형 관계를 학습할 수 있다. 입력층, 은닉층, 출력층으로 구성되며, 각 층의 뉴런이 활성화 함수를 통해 데이터를 학습한다.

3.2. Model Performance Evaluation

회귀 모델의 성능을 평가하기 위하여 평균 절대 오차(MAE), 평균 제곱 오차(MSE), 평균 제곱근 오차(RMSE), 결정 계수(R^2), 평균 절대 백분율 오차(MAPE), 중앙값 절대 오차(MedAE)를 사용하였다.

MAE 는 실제 값과 예측 값 간의 절대 오차의 평균으로 값이 작을수록 예측이 실제에 가까움을 의미한다. MSE 는 오차를 제곱하여 평균한 값으로 예측 정확도를 평가하는 데 중요하다. RMSE 는 MSE 의 제곱근이고, 값이 작을수록 예측의 정확성이 높다. R^2 은 모델이 실제 값을 얼마나 잘 설명하는지를 비율로 나타내며, 1 에 가까울수록 설명력이 높다. MAPE 는 예측 오차를 백분율로 나타내어, 값이 작을수록 예측 정확도가 높음을 의미한다. MedAE 는 예측 오차의 중앙값으로, 작을수록 예측이 일관되며, 큰 오차의 영향을 덜 받는다.

3.2.1. Model Performance Metrics

모델 성능 지표 분석 결과는 다음과 같다.

예측 정확도(R^2)에서 XGBoost 회귀 모델과 랜덤 포레스트 회귀 모델이 각각 0.98 로 가장 높은 설명력을 보였으며, 우수한 예측 성능을 입증했다. 다층 퍼셉트론 회귀 모델 또한 0.98 로 높은 설명력을 나타냈으나, 다른 지표에서는 상대적으로 낮은 성능을 보였다. 반면, 서포트 벡터 회귀 모델은 0.43 으로 가장 낮은 값을 기록하여, 데이터 특성을 설명하는 데 한계가 있다.

오차 크기(MAE, MSE, RMSE)는 XGBoost 회귀 모델(0.99, 3.09, 1.75)과 랜덤 포레스트 회귀 모델(0.99, 3.05, 1.75)이 가장 낮은 오차를 기록하며, 예측이 실제 값에 매우 근접한다. 반면, 선형 회귀 모델(6.40, 82.41, 9.08)과 서포트 벡터 회귀 모델(4.48, 107.18, 10.35)은 가장 높은 오차를 나타냈다.

상대적 오차(MAPE)는 XGBoost 회귀 모델이 15.24%로 가장 낮아 상대적 오차가 적고 예측의 일관성이 뛰어나며, 선형 회귀 모델은 758.06%로 가장 높은 값을 보였다.

중앙값 절대 오차(MedAE)는 XGBoost 회귀 모델(0.37)과 랜덤 포레스트 회귀 모델(0.42)이 가장 낮은 값을 기록하여 안정적인 예측 성능을 보였다. 반면, 선형 회귀 모델(5.22)은 가장 높은 오차 중앙값을 보였다.

그래디언트 부스팅 회귀 모델과 다층 퍼셉트론 회귀 모델은 각 지표에서 보통 수준의 값을 보였으나 XGBoost 회귀 모델과 랜덤 포레스트 회귀 모델에 비해 상대적으로 높거나 낮아 가장 우수한 예측 성능을 가진 모델로 판단하기에는 부족하다.

XGBoost 회귀 모델과 랜덤 포레스트 회귀 모델이 가장 우수한 예측 성능을 보였으며, 특히 XGBoost 회귀 모델이 약 2% 더 낮은 MAPE 값을 기록하여 가장 안정적이고 정확한 예측 성능을 가진 모델로 판단된다.

Table 1: Evaluation Metric

Algorithm	MAE	MSE	RMSE
Linear	6.40	82.41	9.08
Random Forest	0.99	3.05	1.75
Support Vector	4.48	107.18	0.43
Gradient Boosting	3.67	29.70	5.45
XGBoost	0.99	3.09	1.75
Multi-Layer Perceptron	1.16	4.17	2.04

Table 2: Evaluation Metric

Algorithm	R ²	MAPE(%)	MedAE
Linear	0.56	758.06	5.22
Random Forest	0.98	17.83	0.42
Support Vector	10.35	44.01	0.80
Gradient Boosting	0.84	258.96	2.45
XGBoost	0.98	15.24	0.37
Multi-Layer Perceptron	0.98	36.86	0.61

3.3. Final Prediction Model

XGBoost 회귀 알고리즘을 사용하여 예측 모델을 구축하고, 각 질병별로 1인 가구와 다인 가구의 연령대별 발생률 차이를 시각화하였다. 분석 결과, 모든 질병에서 1인 가구가 다인 가구에 비해 전반적으로 높은 발생률을 보이며, 다인 가구는 낮고 안정적인 발생률이 나타난다.

위궤양(Figure 1)은 40대 중반에서 50대 중반에 발생률이 가장 높고, 65세 이후부터는 감소하는 경향을 보인다. 이는 중장년층의 1인 가구가 위궤양 위험이 높음을 시사한다.

위축성 위염(Figure 2)은 40대 후반에서 60대 초반에 발생률이 가장 높으며, 연령 증가에 따른 발생률 상승 폭이 다인 가구보다 훨씬 크다.

장상피화생(Figure 3)은 40대 후반에서 50대 초반에 급격히 증가한 뒤 감소하는 양상을 보인다.

위용종(Figure 4)은 모든 연령대에서 1인 가구의 발생률이 높으며, 연령이 증가할수록 발생률이 상승하다가 60대 이후로는 감소하는 경향이 나타난다.

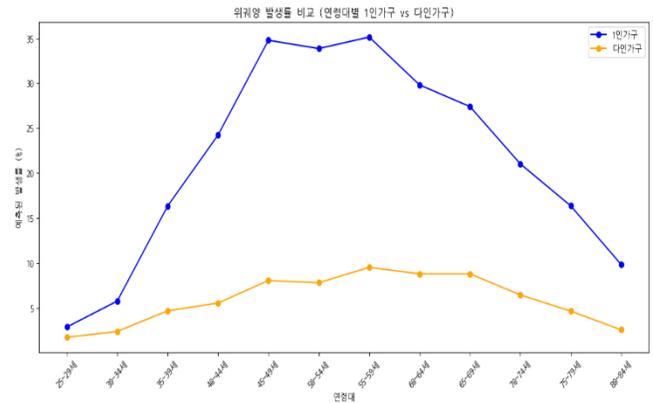


Figure 1: Incidence Rate of Gastric Ulcer

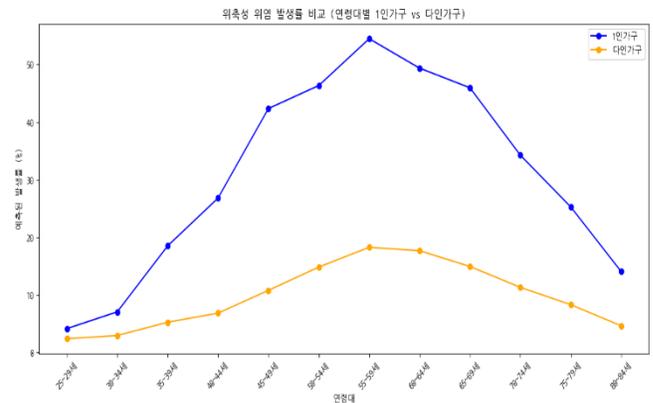


Figure 2: Incidence Rate of Atrophic Gastritis

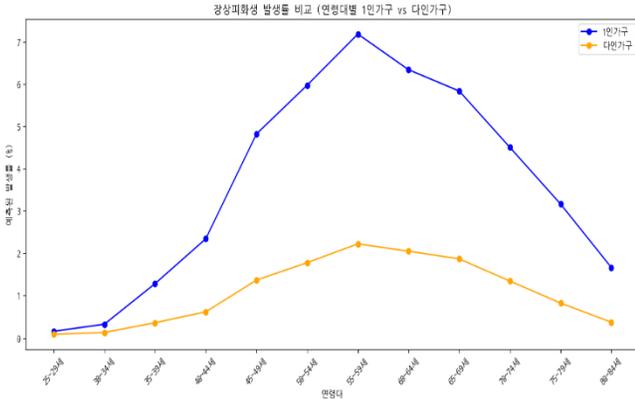


Figure 3: Incidence Rate of Intestinal Metaplasia

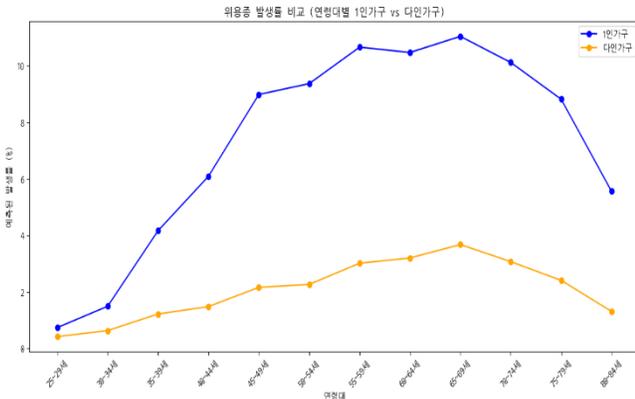


Figure 4: Incidence Rate of Gastric Polyps

4. Conclusion

본 연구에서는 1인 가구가 대부분의 연령대에서 다인 가구보다 위장질환 발생률이 높으며, 특히 중년층에서 발생률 증가가 두드러진다는 결과를 확인하였다. 반면, 20대에서는 발생률이 낮고, 가구 형태에 따른 차이도 크지 않았다. 그러나 30대 중반 이후 발생률이 급격히 상승하는 만큼, 20대의 불규칙한 식습관을 방치해서는 안 된다는 점도 시사한다. 이를 통해 1인 가구가 다인 가구에 비해 위장질환 위험이 높다는 점이 드러났으며, 특히 중장년층 1인 가구를 위한 건강한 식습관 프로그램의 필요성이 강조된다.

이번 연구는 1인 가구 증가와 관련된 식습관 문제에 대한 사회적 관심의 필요성을 제기하며, 1인 가구의 사회적 고립이 식습관 문제를 심화할 가능성이 있음을 시사한다. 특히, 중장년층과 노년층의 결식 문제의 원인은 사회적 고립과 밀접한 관련이 있다. 서울시의 1인 가구 실태 조사에 따르면, 노년층 1인 가구가 혼자 식사하기 싫어 식사를 거른

비율이 다른 세대보다 높다고 한다. 이를 예방하기 위해 노년층 1인 가구에 대한 신체적, 정서적, 사회적 지원과 사회적 교류 촉진 프로그램이 필요하며, 사회 전체가 1인 가구의 건강한 생활과 사회적 고립 완화를 위해 노력해야 한다.

그러나 본 연구에는 몇 가지 한계가 있다. 첫째, 사용된 데이터가 특정 연도를 기반으로 하여 모든 인구를 대표하기 어려운 점이 있다. 둘째, 가구 형태와 연령대 외의 생활 습관, 소득 수준 등의 요인을 고려하지 않아, 위장질환 발생에 영향을 미칠 수 있는 다른 변수들을 포함한 추가 연구가 필요하다.

References

Asan Medical Center. (n.d.). Chronic gastritis. Retrieved from <https://cancer.amc.seoul.kr/asan/healthinfo/disease/diseaseDetail.do?contentId=31402>

Asan Medical Center. (n.d.). Gastric polyps. Retrieved from <https://www.amc.seoul.kr/asan/healthinfo/disease/diseaseDetail.do?contentId=32058>

Asan Medical Center. (n.d.). Peptic ulcer. Retrieved from <https://www.amc.seoul.kr/asan/healthinfo/disease/diseaseDetail.do?contentId=31621>

Cho, H. J. (2024). Changing eating habits can prevent gastric cancer. *Korean Medical Communication*. Retrieved from <https://www.kmpnews.co.kr/news/articleView.html?idxno=60904>

Do, K., & Shim, Y. (2023). Comparison of eating habits and dietary behaviors of young single-person households by dietary lifestyle. *Korean Journal of Health Promotion*, 2234(2141), 2093-5676.

Heo, Y. K., & Sim, K. H. (2016). Dietary attitudes of single-person households in metropolitan areas. *The Korean Journal of Food and Nutrition*, 29(5), 735-745.

Hadong County Public Health Center. (n.d.). Gastrointestinal disorders. Retrieved from <https://www.hadong.go.kr/01868/01959/02057.web>

Kim, E. K., Kwon, Y. S., Lee, J. Y., Park, Y. H., Jang, H. J., & Kim, D. S. (2022). Needs assessment for food education and support programs for young single-person households in the Seoul metropolitan area. *Journal of the Korean Society of Food Culture*, 37(6).

Korea Rural Economic Institute. (2023). *2023 Food consumption behavior survey: Statistical report*. Retrieved from <https://www.krei.re.kr/foodSurvey/selectBbsNttView.do?key=806&bbsNo=449&nttNo=164071>

Lee, M. Y. (2017). Eating alone increases risks of obesity and gastrointestinal disorders. *Sisa Economy News*. Retrieved from <https://www.sisanews.kr/news/articleView.html?idxno=23614>

Moon, S. H. (2022). Atrophic gastritis and intestinal metaplasia that are difficult to reverse: Endoscopy as a solution. *Medical Times*. Retrieved from <https://www.medicaltimes.com/Main/News/NewsView.html?I>

D=1148986

Seoul Institute. (2022). *A study on the actual status of single-person households in Seoul*. Seoul Metropolitan Government Single-Person Household Special Measures Promotion Task Force.

Statistics Korea. (2024). Households by head's gender, age, and household composition (general households) - Si/Gun/Gu. Retrieved from https://kosis.kr/statHtml/statHtml.do?orgId=101&tblId=DT_1JC1517

Statistics Korea. (2024). Projections of future households: 2022~2052. Retrieved from https://kostat.go.kr/board.es?mid=a10301020600&bid=207&act=view&list_no=432696

Yoon, T. S., & Min, J. H. (2023). Factors influencing depression in single-person households by age group: A comparison with multi-person households. *Journal of Family and Quality of Life Studies*, 41(1), 77-91.