

회귀분석 프로젝트

OECD 국가별 고용률 요인 분석



과목명	회귀분석
학과	데이터사이언스학과
학번	12211262
성명	우지민

목 차

1. 개요 및 목적	4
1.1 연구목적 및 연구질문	4
1.2 데이터 설명	4
2. 방법론.....	5
2.1 변수	5
2.1.1 반응변수	5
2.1.2 설명변수	5
2.2 분석 방법론	7
3. 탐색적 데이터 분석.....	8
3.1 변수 요약	8
3.2 변수 시각화	9
3.2.1 단변량 분석	9
3.2.2 이변량 분석	11
4. 변수 선택	13
4.1 VIF	13
5. 분석 결과	14
5.1 모델 적합	14
5.1.1 R-squared	14
5.1.2 F-statistics	14
5.2 계수 분석	14
6. 회귀진단	15
6.1 가정 확인	15
6.1.1 잔차 정규성	15

6.1.2 등분산성	15
6.1.3 잔차 독립성	16
6.1.4 영향점 진단	16
6.2 데이터 변환	17
6.2.1 1 차 차분	17
7. 결론.....	17
7.1 해석	17
7.2 한계	19
8. 참고문헌	19

1. 개요 및 목적

1.1 연구목적 및 연구질문

고용률(Employment to population ratio)은 한 국가의 노동시장 성과의 경제활력도를 평가하는 핵심 지표이다. 국제기구(OECD, World Bank, UNDP 등)는 고용률이 경제성장, 교육수준, 기술투자, 산업구조, 인구구성 등 복합적인 요인에 의해 결정된다고 보고하고 있다. 또한 최근 인구 고령화, 글로벌 경기변동 등 경제적 불확실성이 심화되며, 어떤 요인이 고용률 향상에 기여하는지 규명하는 것은 정책적, 학술적으로 중요한 의미를 갖는다. 본 연구는 OECD 회원국들의 데이터를 바탕으로 경제, 사회 지표들이 고용률에 미치는 영향을 실증적으로 분석하고자 한다.

본 보고서는 다음의 핵심 질문을 탐구한다.

RQ1. 어떤 경제, 사회적 요인이 국가의 고용률 상승 또는 하락을 유의하게 설명하는가?

RQ2. 경제적 변수(GDP, CPI, FDI 등)와 사회적 변수(교육, 인구구조, 노동, 제도 등)는 고용률에 어떻게 다른 영향을 미치는가?

1.2 데이터 설명

본 연구는 다음의 국제기구 공공데이터를 활용하였다.

데이터 출처	데이터
OECD Data Explorer	고용률, 시간당 노동생산성, 연간 근로시간, 연구개발 지출, 외국인 인구 유입, 고용보호 엄격성 지수 등
World Bank Group Data	1인당 GDP, 소비자물가지수, 제조업 부가가치, 서비스업 부가가치, 외국인직접투자 비율, 65세 이상 인구 비중, 합계출산율, 도시 인구 비중, 여성 노동참가비율, 출생 시 기대수명, 지니계수, 인터넷 이용자 비율, 연구개발 지출, 일반정부 최종소비지출 등
UNDP Human Development Reports	기대교육연수

분석대상 국가는 OECD 회원국 41개 국가, 분석기간은 2020-2023년이며 데이터의 가용성을 고려하여 국가 목록과 년도 범위가 선택되었다.

2. 방법론

2.1 변수

2.1.1 반응변수

변수 분류	변수명	출처	설명	단위
반응변수	Employment_to_population_ratio	OECD	고용률. 전체 인구 중 취업자가 차지하는 비율.	전체 인구 대비 백분율 (%)

2.1.2 설명변수

국가의 고용률에 영향을 미칠 수 있는 경제·사회·제도 지표들은 기존 노동경제학 및 거시경제 분야의 대표적 이론에 근거하여 선정되었다. 특히 Solow 성장모형[1], Lewis 이중경제모형[2], 인적자본 이론[3], 내생적 성장모형[4] 등 고용과 경제성장의 상호작용을 설명하는 핵심 이론들을 참고하였다.

변수 분류	변수명	출처	설명	단위
경제 변수	1. GDP_per_capita	WB	1인당 GDP. 한 국가에서 일정 기간 동안 생산된 재화·서비스의 총소득을 인구수로 나눈 값.	미 달러(USD)
	2. GDP_growth	WB	국내총생산(GDP)의 전년 대비 실질 변화율.	연간 증가율 (%)
	3. Inflation_consumer_prices	WB	소비자물가지수(CPI)로 측정한 물가상승률. 일정 기간동안 소비자가 구매하는 상품·서비스 묶음 가격의 평균 변동을 나타냄.	연간 증가율 (%)
	4. Manufacturing_value_added	WB	제조업 부가가치. 생산된 총산출물에서 중간투입물(재화·서비스)을 제외한 순가치로, 제조업이 경제에 기여하는 정도.	GDP 대비 비율 (%)
	5. Services_value_added	WB	서비스업 부가가치. 서비스산업이 생산활동을 통해 창출한 부가가치.	GDP 대비 비율 (%)
	6. GDP_per_hour_worked	OECD	시간당 노동생산성. 노동자 한 명이 시간당 창출한 GDP.	PPP 기준 시간당 미 달러(USD)

	7. Foreign_direct_investment_net_inflows	WB	외국인직접투자(FDI) 순유입. 해외 투자자가 해당 국가에 유입한 순자본.	GDP 대비 비율 (%)
인구·사회 변수	8. Population_ages_65_and_above	WB	65 세 이상 인구 비중.	전체 인구 대비 비율 (%)
	9. Fertility_rate_total	WB	합계출산율. 여성 1 명이 평생 낳을 것으로 예상되는 평균 출생아 수.	여성 1 인당 출생 수
	10. Urban_population	WB	도시 인구 비중. 도시 지역에 거주하는 인구의 비율.	전체 인구 대비 비율 (%)
	11. Expected_Years_of_Schooling	UNDP	기대교육연수. 평균적으로 한 개인이 평생 받을 것으로 예상되는 교육 연수.	년(Years)
	12. Labor_force_participation_rate_female	WB	여성 노동참가율(15~64 세). 여성 인구 중 경제활동에 참여하는 비율.	여성 인구 대비 비율 (%)
	13. Inflows_of_foreign_population	OECD	외국인 인구 유입(이민 유입자 수).	총 유입 인구(명)
	14. Life_expectancy_at_birth_total	WB	출생 시 기대수명. 현재의 사망률 기준으로 개인이 생존할 것으로 기대되는 평균 연령.	년(Years)
	15. Gini_index	WB	지니계수. 소득 또는 소비지출 분배의 불평등 정도를 나타내는 지표(0: 완전평등, 100: 완전불평등).	0~100 점수
	16. Individuals_using_the_Internet	WB	인터넷 이용자 비율. 일정 기간 내 인터넷을 사용한 인구 비율.	전체 인구 대비 비율 (%)
노동·제도 변수	17. Individual_and_collective_dismissals	OECD	고용보호 엄격성 지수. 정규직·비정규직 해고 규제, 단체해고 요건 등의 엄격성을 평가한 지수.	0~6 점수
	18. Hours_worked	OECD	연간 1 인당 실제 근로시간. 노동자가 연간 실제로 일한 총 시간.	연간 근로시간(시간/년)
	19. Research_and_development_expenditure	WB	연구개발(R&D) 지출. 연구 및 기술혁신 활동에 투입된 지출 규모.	GDP 대비 비율 (%)

20. General_government_financial_consumption_expenditure	WB	일반정부 최종소비지출. 공공서비스 제공을 위해 정부가 지출한 소비 비용.	GDP 대비 비율 (%)
--	----	--	---------------

2.2 분석 방법론

본 연구는 다음의 절차에 따라 분석을 진행하였다.

1) 데이터 전처리

서로 다른 출처의 데이터들을 Country와 Year 기준으로 병합하였고, 기존 OECD 회원국 52개 국가 중 결측값이 많은 일부 국가들 ('ARG', 'BRA', 'CYP', 'IND', 'MKD', 'MLT', 'PER', 'PRY', 'RUS', 'URY', 'ZAF')의 데이터는 제외하여 총 41개 국가로 범위를 축소하였다. 수집년도의 경우에도 결측값을 고려했을 때 유의미한 분석을 위해, 2020~2023년의 4개년으로 기간을 한정하였다. 그 외의 결측치는 각 열별 중앙값으로 대체하였다.

2) 다중공선성 검토

모든 설명변수에 대해 Variance Inflation Factor(VIF)을 계산하였고, VIF > 10인 변수를 반복적으로 제거하여 안정적 모형을 확보하고자 하였다. 최종적으로 8개의 설명변수를 유지하였다.

3) 다중회귀분석

Ordinary Least Squares 방법을 활용하였고, 회귀모델 식은 아래와 같다.

$$Employment = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon$$

4) 회귀진단 수행

아래와 같은 절차를 통해 모델의 가정을 확인하였다.

- 가) 선형성 확인 : Residuals vs. Fitted plot
- 나) 정규성 확인 : Q-Q plot
- 다) 등분산성 확인 : Residuals vs. Standardized Residuals
- 라) Autocorrelation 확인 : Durbin-Watson test
- 마) 영향점 확인 : Cook's distance

5) 최종 결과 해석 및 개선

분석 결과를 바탕으로 추정된 회귀계수들의 방향과 크기를 해석하고, p-value 기반 통계적 유의성을 평가하였다. 또한 추가적으로 변수 변환을 통해 적합 결과를 개선하고자 하였다.

3. 탐색적 데이터 분석

3.1 변수 요약

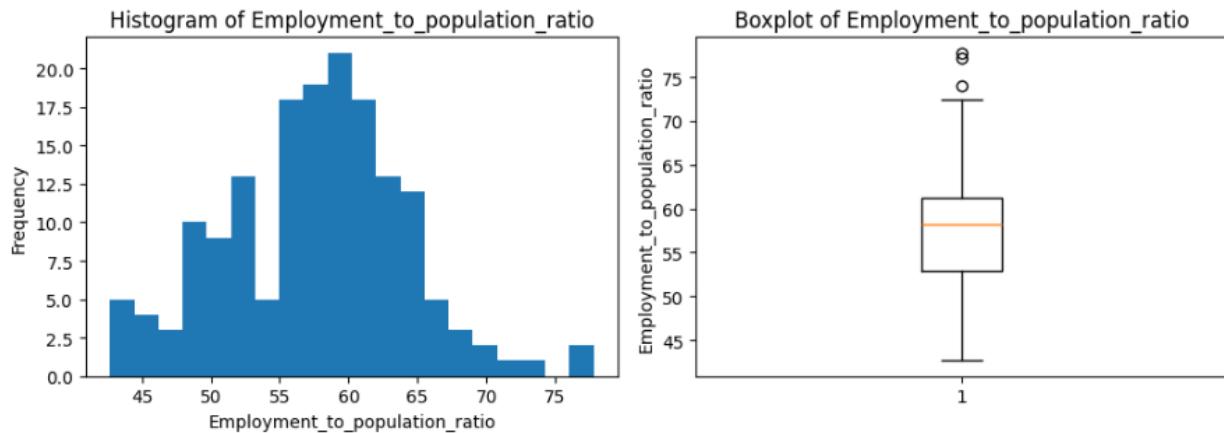
본 연구에서 활용한 주요 경제, 사회 변수에 대해 기초 통계량(평균, 중앙값, 표준편차, 최소값, 최빈값)을 산출하였다. 1인당 GDP의 최소값과 최대값을 보면 격차가 크게 나타났으며, OECD 국가들 사이에도 소득수준 차이가 매우 큼을 확인할 수 있었다. 또한 일부 변수들은 양(+)과 음(-)의 비율 변화를 보이고 있음을 확인할 수 있었다.

		Mean	Median	Standard Deviation	Minimum	Maximum
	Year	2021.50	2021.50	1.12	2020.00	2023.00
	Employment_to_population_ratio	57.58	58.26	6.58	42.69	77.80
	GDP_per_hour_worked	62.72	55.49	24.15	18.33	142.90
	Inflows_of_foreign_population	203405.27	95711.50	317582.85	1782.00	2481019.00
	Hours_worked	1649.55	1620.60	208.54	1207.35	2287.54
	GDP_per_capita	41852.23	34595.58	27449.87	5339.69	134965.82
	GDP_growth	1.99	2.50	4.82	-10.94	16.26
	Inflation_consumer_prices	5.83	3.91	7.82	-1.25	72.31
	Manufacturing_value_added	14.03	12.79	5.58	3.88	37.15
	Services_value_added	63.88	63.39	6.52	41.82	81.13
	Foreign_direct_investment_net_inflows	0.25	2.81	32.62	-391.56	105.64
	Population_ages_65_and_above	18.48	19.54	4.34	7.49	29.56
	Fertility_rate_total	1.53	1.52	0.31	0.72	3.00
	Urban_population	77.75	81.15	11.69	53.76	98.19
	Labor_force_participation_rate_female	69.34	71.73	9.65	34.74	84.05
	Life_expectancy_at_birth_total	79.96	81.29	3.31	69.75	84.56
	Gini_index	33.13	32.40	6.01	23.80	55.10
	Individuals_using_the_Internet	89.40	91.00	7.06	69.80	99.83
	Research_and_development_expenditure	1.92	1.70	1.06	0.26	6.02
	General_government_final_consumption_expenditure	19.71	20.49	3.84	11.06	28.07
	Expected_Years_of_Schooling	17.00	16.62	1.73	14.05	21.08

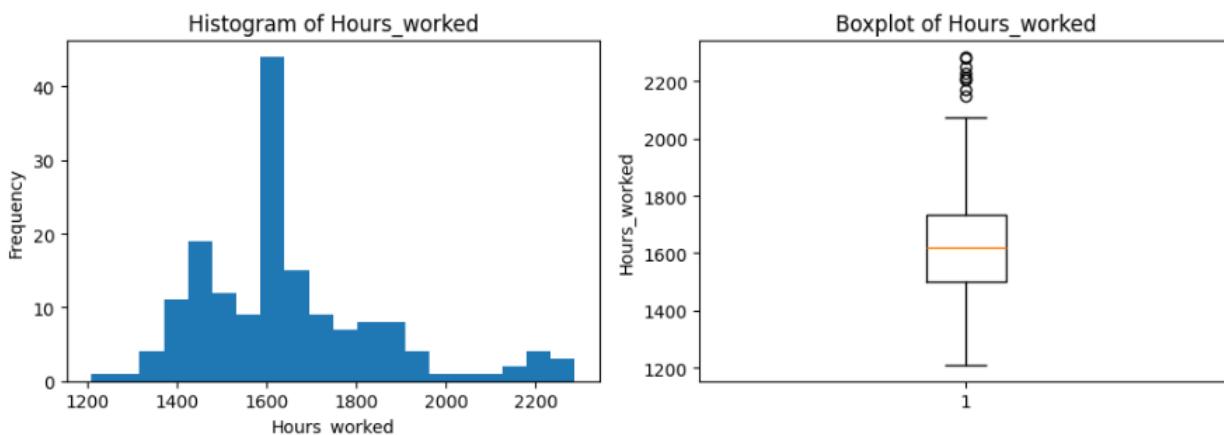
3.2 변수 시각화

3.2.1 단변량 분석

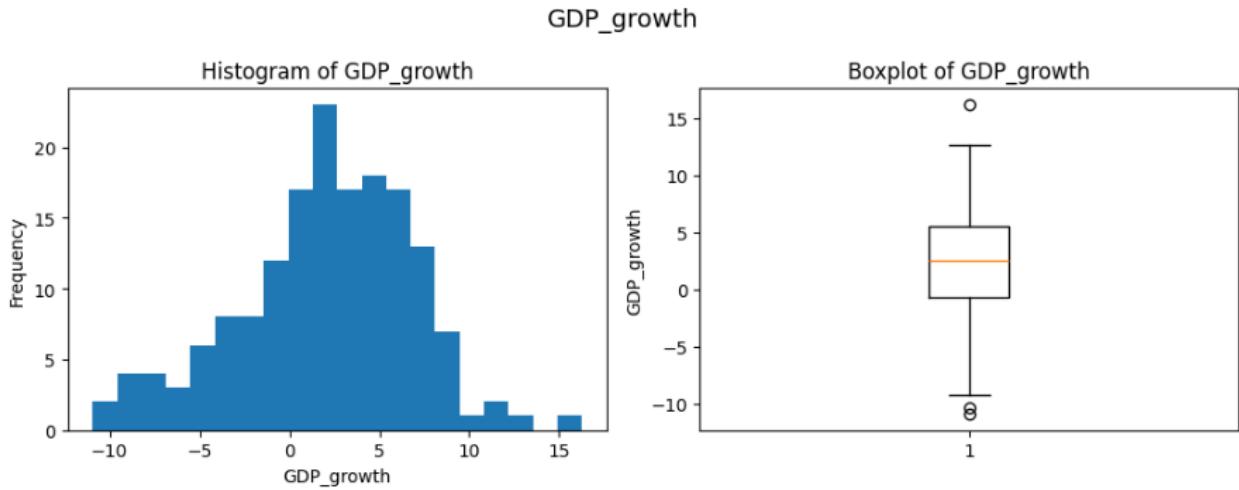
히스토그램과 박스플롯으로 아래와 같이 각 변수의 분포를 확인하였다.



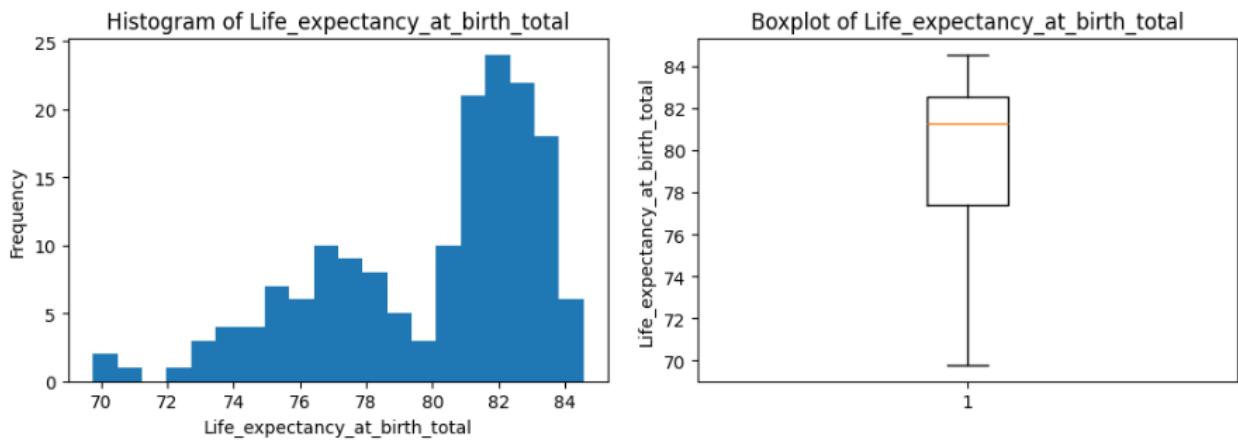
대부분의 국가의 고용률은 55~65% 범위에 몰려 있으며, 고용률이 매우 높은 국가(70%대)가 일부 존재하는 것으로 확인되었다. 노동자 한 명이 시간당 창출한 GDP는 40~60 달러와 80~90 달러로 나뉘어 집중되는 분포를 보였고, 이는 국가들 사이에 비교적 뚜렷한 차이가 존재하는 것으로 해석된다. 외국인 이민 유입자 수는 0~25만 명 범위에 몰려 있고, right-skewed 분포를 보이는 것으로 나타났으며, 100만 명 이상의 이민 유입자 수가 매우 높은 국가들도 일부 존재하는 것으로 확인되었다.



연간 1인당 근로 시간은 1600~1700시간 범위에 몰려 있으며, right-skewed 분포를 보이는 것으로 나타났다. 1인당 GDP의 경우, 2~5만 달러 사이에 몰려 있고, right-skewed 분포를 보이는 것으로 나타났다.

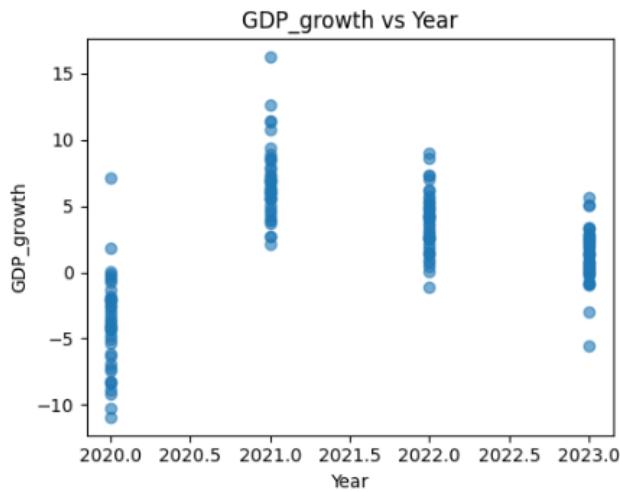


국내총생산(GDP)의 전년 대비 실질 변화율(GDP 성장률)은 양(+)과 음(-)의 퍼센트가 모두 존재하며 약 2%가 평균으로 나타났고, left-skewed 분포를 보였다. 소비자 물가지수 증가율은 0~10% 범위에 몰려 있는 것으로 나타났다. 제조업 부가가치 비율(GDP 대비)은 10~20% 범위에 몰려 있으며, right-skewed 분포를 보이는 것으로 나타났다. 서비스업 부가가치 비율(GDP 대비)은 55~70% 범위에 몰려 있는 것으로 나타났다. 외국인직접투자(FDI) 순유입 비율(GDP 대비)은 중앙값이 2.81%인 것으로 나타났고, 0% 주위로 몰려있는데 일부 극소수의 국가에 대해 매우 큰 FDI 양(+)과 음(-)의 퍼센트값이 나타났다. 전체 인구 대비 65세 이상 인구의 비중(%)은 20% 주위로 몰려 있는 것으로 확인되었으며, left-skewed 분포를 보이는 것으로 나타났고, 일부 소수의 국가에 대해 30%에 근접한 값을 보이는 것으로 나타났다. 이는 고령화의 정도가 매우 높은 일부 국가들에 해당하는 것이다. 합계출산율의 경우에는 평균이 1.53명으로 그래프의 형태가 거의 symmetric하게 나타났다. 다만, 합계출산율이 1.0보다 낮거나, 3.0에 근접한 일부 소수의 국가들도 box plot에서 이상치로 표현되었다. 전체 인구 대비 도시 인구 비중은 60% 미만부터 100%에 근접한 국가들까지 비교적 고르게 분포하였고, 중앙값은 81.5%로 나타났다. 여성 인구 대비 노동참가율(15~64세, %)은 70~80% 범위에 몰려 있으며, left-skewed 분포를 보이는 것으로 나타났다.

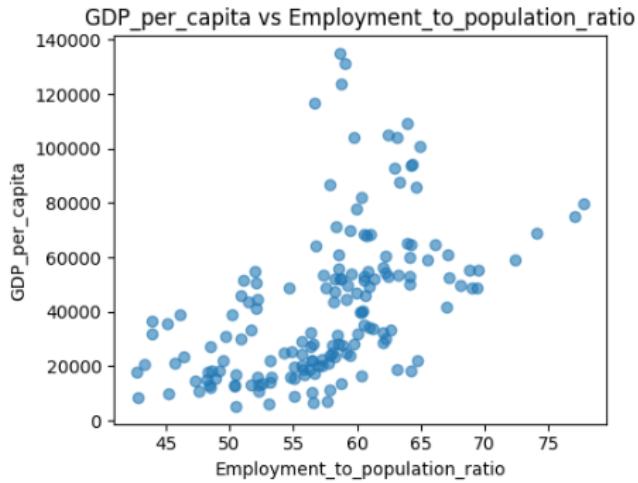


출생 시 기대수명의 경우에는 76~78세와 82세 주위로 나뉘어 분포 형태가 나타났고, 이는 국가들 사이에 비교적 뚜렷한 차이가 존재하는 것으로 해석되며 전반적으로 강하게 left-skewed 분포를 보였다. 소득 또는 소비지출 분배의 불평등 정도를 나타내는 지표(0~100)인 지니 계수의 경우에는 30~35 사이에 몰려 있으며, right-skewed 분포로 나타났다. 전체 인구 대비 인터넷 이용자 비율은 85~100% 사이에 고르게 분포하였고, 일부 70~75% 범위의 국가들도 존재하는 것으로 나타났다. 연구개발(R&D) 지출(GDP 대비 비율, %)의 경우 1%대 후반에 몰려 있고, 일반정부 최종소비지출(GDP 대비 비율, %)의 경우 20~22.5% 사이에 몰려 있는 것으로 나타났다. 기대교육연수의 경우, 중앙값은 16.62년이고, 14년부터 21년까지 비교적 고르게 분포하는 것으로 나타났다.

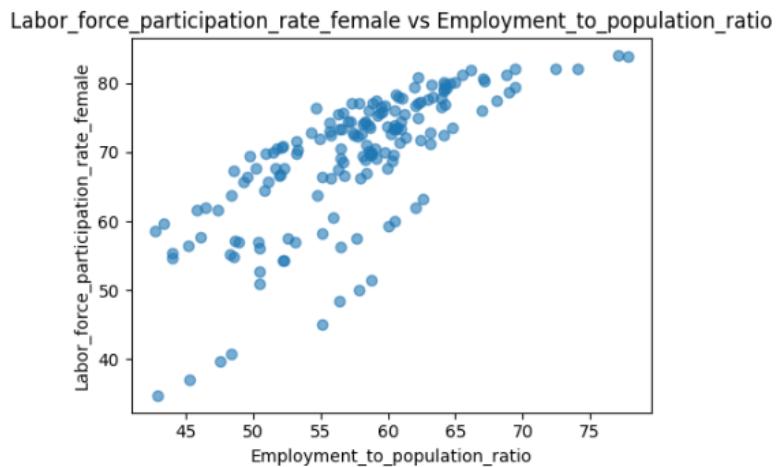
3.2.2 이변량 분석



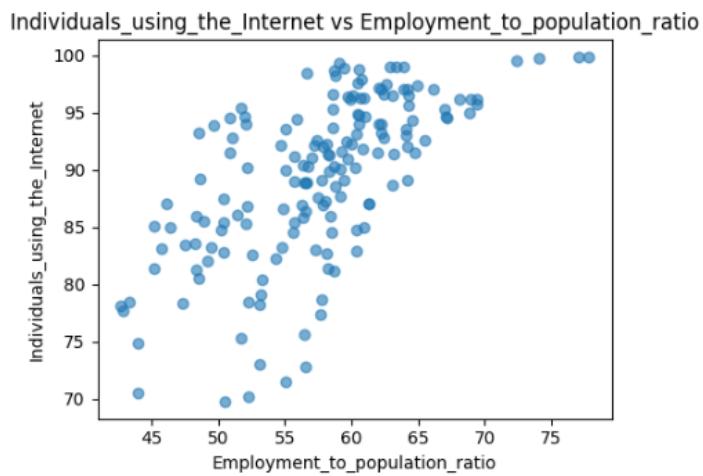
GDP_growth vs Year 산점도 분포를 통해 GDP 성장률은 2020년에는 대부분의 국가에 대해 음(-)의 방향 성장률이었고, 2021년에는 양(+)의 방향 추세로 바뀌었으며, 2022, 2023년은 저 성장 추세로 변하는 것으로 나타났다. 이는 코로나19로 인한 세계적 경기 침체와 반등효과 등으로 인해 나타난 추세라고 추정할 수 있다.



GDP_per_capita vs Employment_to_population_ratio 산점도 분포를 통해 전체적으로 명확한 양(+)의 선형 관계가 관찰되었다. 따라서 고용률이 높은 국가일수록 GDP per capita도 높은 경향을 보였다. 고용률이 60% 이상인 국가들의 GDP per capita는 대체로 3~10만 달러 이상으로 나타나며, 반면 고용률이 45~55% 수준인 국가들은 GDP per capita가 상대적으로 낮은 1~4만 달러대가 많은 것으로 나타났다.



Labor_force_participation_rate_female vs Employment_to_population_ratio 산점도 분포를 통해 전체적으로 매우 명확한 양(+)의 선형 관계가 관찰되었다. 따라서 여성의 노동시장 참여 확대는 국가 전체의 고용률을 높이는 핵심 요인으로 추정할 수 있다. 여성 노동참여율이 70~80% 이상인 국가들은 고용률도 대부분 55~75% 수준으로 높은 것으로 나타났다. 반면, 45~55% 수준의 고용률이 낮은 국가들은 여성 노동 참가율 역시 35~60% 수준으로 낮은 것으로 나타났다.



`Individuals_using_the_Internet` vs `Employment_to_population_ratio` 산점도 분포를 통해 전체적으로 완만한 양(+)의 선형 관계가 존재함을 확인할 수 있었다. 따라서 전반적으로 정보접근성, 디지털화 수준이 높은 국가일수록 고용률이 높은 구조가 나타나는 경향을 확인할 수 있었다. 인터넷 이용률이 높은 국가(95~100%)는 고용률도 대체로 55~70%로 높은 편이고, 반면 인터넷 이용률이 70~85% 수준인 일부 국가에서는 고용률도 상대적으로 낮은 경향을 확인할 수 있었다.

4. 변수 선택

4.1 VIF

다중공선성(Multicollinearity)을 제거하기 위해 각 설명변수에 대해 Variance Inflation Factor(VIF)를 계산하여 반복적으로 $VIF > 10$ 기준에 따라 변수를 제거한 결과, 설명변수 후보 20개 중 12개의 변수가 제거되어 최종적으로 남은 설명변수의 개수는 8개 ('Inflows_of_foreign_population', 'GDP_per_capita', 'GDP_growth', 'Inflation_consumer_prices', 'Manufacturing_value_added', 'Foreign_direct_investment_net_inflows', 'Gini_index', 'Research_and_development_expenditure')가 되었다. 최종 8개 변수는 모두 $VIF < 10$ 으로 회귀계수의 신뢰성과 해석 가능성을 확보하였다.

5. 분석 결과

5.1 모델 적합

5.1.1 R-squared

최종 8개의 설명변수에 대해 OLS 방식으로 적합한 결과, R-squared = 0.325, Adjusted R-squared = 0.290으로 나타났다. 따라서 모형은 국가 간 고용률 변동의 약 29~33%를 설명하는 것으로 해석할 수 있다. 거시경제 및 인구구조가 복합적으로 작용하는 특성을 고려한다면 중간 수준의 설명력이라고 할 수 있다.

5.1.2 F-statistics

최종 8개의 설명변수에 대해 OLS 방식으로 적합한 결과, F-statistic = 9.310, 전체모형의 유의 확률 p-value = 1.93e-10으로 나타났다. 따라서 1% 유의수준에서 모형 전체가 통계적으로 유의하며, 선택된 설명변수들은 집합적으로 고용률을 설명하는 데 의미 있는 영향력을 갖는다고 해석할 수 있다.

5.2 회귀계수 분석

회귀계수, t-value, p-value를 바탕으로 각 설명변수가 고용률에 미치는 방향성과 유의성을 확인하였다. p-value < 0.05으로 통계적으로 유의한 것으로 판단한 변수는 아래와 같다.

가) 1인당 GDP (GDP_per_capita)

Coefficient = 0.0001, t = 6.464, p < 0.001의 결과를 보였다. 따라서 1인당 GDP가 증가할수록 고용률이 유의하게 증가하는 양(+)의 관계로 볼 수 있다. 국가의 1인당 GDP가 10,000 달러 증가할 때, 고용률은 평균적으로 약 1% 포인트 상승한다고 해석할 수 있다.

나) R&D 지출 비중 (Research_and_development_expenditure)

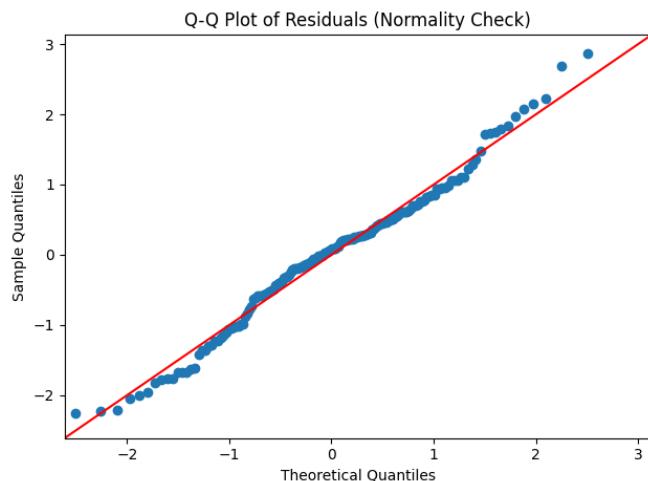
Coefficient = 1.0802, t = 2.440, p < 0.05의 결과를 보였다. 따라서 R&D 지출 비중이 1%p 증가하면 고용률이 약 1.08%p 증가한다고 해석할 수 있다. 이는 기술혁신 및 지식기반 산업이 성장할수록 더 많은 고용기회가 창출될 수 있음을 의미한다고 해석할 수 있다.

6. 회귀진단

6.1 가정 확인

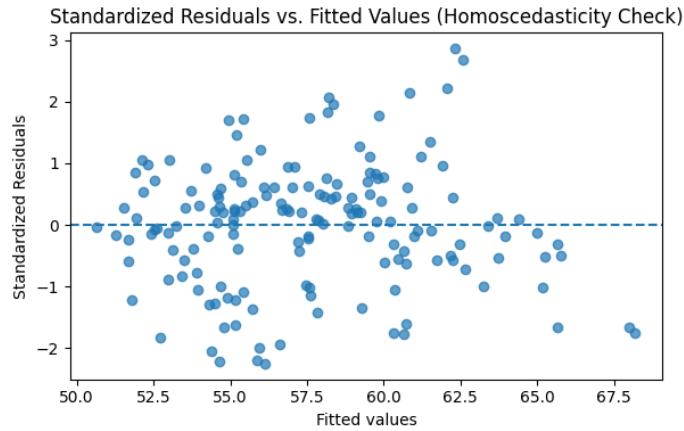
6.1.1 잔차 정규성

Q-Q Plot (Quantile-Quantile Plot)은 회귀모형의 잔차가 정규분포를 따르는지를 시각적으로 평가하기 위한 방법이다. 아래 Q-Q Plot 결과에 따르면, 대부분의 잔차 점들이 대각선의 기준선을 따라 비교적 고르게 분포하고 있다. 이는 잔차가 정규분포를 따른다는 가정을 전반적으로 충족하고 있음을 의미한다. 다만, 상위 분위수(우측 상단)에서 일부 점들이 기준선에서 약간 벗어나는 경향을 보였는데 그 정도가 크지는 않아 회귀모델의 정규성 가정에 심각한 위배는 아니라고 판단하였다.



6.1.2 등분산성

회귀모형의 등분산성 가정은 모든 수준의 예측값에서 잔차의 분산이 일정해야 한다는 조건이다. 이를 평가하기 위해 표준화 잔차(Standardized Residuals)와 적합값(Fitted Values)의 산점도를 확인하였다. 산점도 결과, 잔차들은 적합값의 전 범위에서 특정한 패턴 없이 무작위로 흩어져 있는 형태를 보였다. 이는 전반적으로 등분산성 가정이 충족되는 것으로 판단하였다.

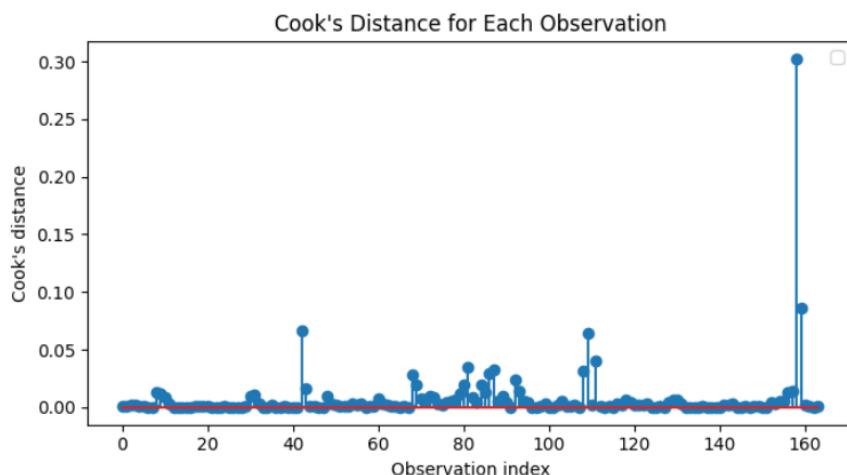


6.1.3 잔차 독립성

Auto-correlation(잔차 자기상관)을 확인하기 위해 Durbin-Watson test를 진행하였고, Durbin-Watson statistic = 0.7887의 결과가 나타났다. 이는 매우 강한 양의 자기상관을 의미하며, 이는 'Year'와 같은 설명변수로 인해 잔차가 시간 방향으로 의존성을 갖게 되었기 때문일 수 있다. 자기상관 문제를 해결하기 위해 '6.2 데이터 변환'에서 1차 차분을 시도해보았다.

6.1.4 영향점 진단

Cook's Distance는 각 관측치가 회귀식 전체에 미치는 영향을 측정하는 지표로, 영향점을 진단할 수 있다. Cook's Distance 그래프를 확인해보면, 모든 포인트가 0.1보다 작은 값을 가지지만 한 점(Index 158)이 다른 모든 점들과 비교했을 때 큰 Distance값을 가지는 것으로 보여, 해당 점을 제거 후 다시 OLS 적합을 진행하였다. 그 결과, 기존 모델과 비교했을 때, R-squared값이 0.3245에서 0.3279로 변하면서 일부 설명력이 높아졌고, 'Inflation_consumer_prices'의 coefficient값이 0.0355에서 0.1287로 일부 커지는 등 회귀계수의 변화를 확인할 수 있었다.

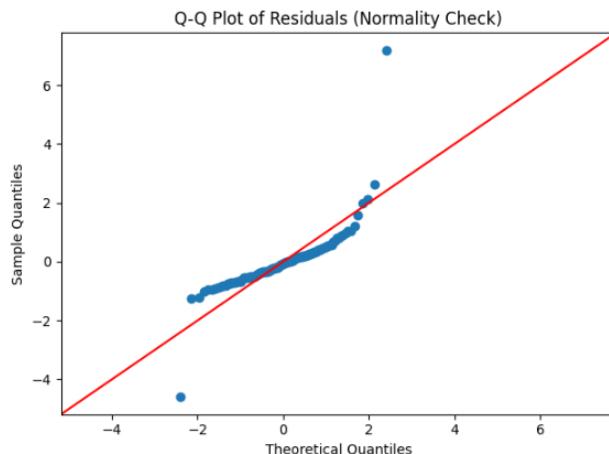


6.2 데이터 변환

6.2.1 1차 차분

패널 데이터(국가, 연도)의 시간에 따른 자기상관(Auto-Correlation)을 제거하기 위해 데이터 변환 방법 중 1차 차분(differencing) 방법을 사용하였다. 변수의 절대수준 대신 연도(Year) 간 변화량을 사용하면 시간 패턴이 제거되어 자기상관이 크게 감소할 수 있기 때문이다.

1차 차분한 데이터(df_diff)를 토대로 OLS 분석을 진행한 결과, Durbin-Watson statistic = 1.75 의 결과가 나타나면서 자기상관성은 완화할 수 있었다. 또한 R-squared = 0.331, Adj R-squared = 0.208, F-statistic = 2.682, p-value = 0.000769로 나타나면서 모형의 설명력 자체는 기존 모델과 유사하거나 다소 감소하였으며, 전체 모형은 통계적으로 유의미한 것으로 나타났다.



다만, Q-Q plot 결과 잔차의 정규성 가정에는 위배되고, $p\text{-value} < 0.05$ 의 유의미한 설명변수의 수가 1개(Labor_force_participation_rate_female)로만 나타났다. 이는 차분 과정에서 특정 연도 충격 등 극단적 변화값이 잔차에 그대로 반영되어 정규성 가정을 충족하지 못하였고, 차분 모형의 특성 상 각 지표의 절대수준이 아니라 연도 간 변동만 반영하기 때문에 많은 변수의 설명력이 감소한 것으로 해석된다.

7. 결론

7.1 해석

본 연구는 OECD, World Bank, UNDP 등 주요 국제지표를 활용하여 국가 간 고용률 수준 및 고용률 변화에 영향을 미치는 경제·사회 요인을 분석하는 것을 목적으로 하였다. 분석은 ①

기존 데이터 기반 OLS 회귀 ② 1차 차분 기반 변화량 분석 두 단계로 수행되었다.

가) 기존 데이터 기반 결과 해석

기존 데이터를 이용한 OLS 회귀모형의 설명력은 R-squared = 0.325, Adjusted R-squared = 0.290으로 나타났다. 이는 모델이 국가 간 고용률 수준의 약 29~33%를 설명하는 것으로, 거시경제·인구구조적 변수가 복합적으로 작용하는 맥락을 고려하면 중간 정도의 설명력으로 해석할 수 있다.

또한, F-statistic의 유의확률이 1.93e-10으로 매우 낮게 나타나, 선택된 설명변수들이 집합적으로 고용률 수준을 유의하게 설명한다는 점이 확인되었다. 계수 분석 결과, 1인당 GDP와 연구개발 지출 비중이 고용률에 통계적으로 유의미한 양(+)의 영향을 주는 변수로 나타났다.

- 1인당 GDP가 10,000달러 증가하면 고용률은 약 1%p 증가
- R&D 지출 비중이 1%p 증가하면 고용률은 약 1.08%p 증가

이는 경제 수준과 기술혁신 투자가 고용시장에 긍정적 영향을 미치는 메커니즘과 일치하며, 지식기반경제로의 전환이 노동시장 활성화로 이어질 수 있음을 시사한다.

다만, 기존 데이터 기반 회귀에서는 Durbin-Watson 통계량(0.7887)이 매우 낮게 나타나, 잔차가 강한 양의 자기상관을 보임이 확인되었다. 이는 국가별 연도 데이터를 포함하는 패널 데이터 특성에서 기인한 것으로 해석되며, 자기상관을 완화하기 위해 1차 차분을 적용하였다.

나) 1차 차분 적용 후 결과 해석

1차 차분을 적용하여 국가별 연도 간 변화량을 분석한 결과, 모형의 설명력은 R-squared = 0.331, Adjusted R-squared = 0.208으로 나타났다. 수준 데이터보다 Adjusted R-squared의 설명력이 다소 낮아졌는데, 이는 차분 기법이 각 변수의 구조적 수준을 제거하고 단기 변동성만을 반영하기 때문으로 이해된다.

차분 모형에서 통계적으로 유의미한 변수는 여성 노동참여율 변화량 (Labor_force_participation_rate_female) 단 하나였다. 여성 노동참여율이 1%p 증가할 때 고용률 변화는 평균적으로 0.41%p 증가하는 것으로 나타났다. 이는 단기적으로 여성의 경제활동 참여가 고용률 변화를 주도하는 주요 요인임을 보여준다.

한편, 차분을 통해 Durbin-Watson 값이 1.75로 개선되며, 잔차 자기상관 문제가 상당 부분 해소된 것이 확인되었다. 다만 일부 변수의 잔차 분포는 여전히 정규성을 만족하지 못하는

한계가 있었다.

7.2 한계

본 연구는 국제기구에서 제공하는 다수의 경제·사회지표를 활용하여 국가 간 고용률 수준 및 변화의 결정요인을 분석하였다는데 의의가 있다. 그러나 다음과 같은 한계가 존재하며, 이는 향후 연구에서 보완할 필요가 있다.

첫째, 본 연구에서 활용한 데이터는 여러 국가의 연도별 정보를 포함한 패널 데이터임에도 불구하고, 분석에는 기본적인 OLS 모형을 적용하였다. 이로 인해 국가별 고유한 특성(정치·제도적 환경, 사회구조, 문화적 요인 등)을 통제하지 못했고, 이는 잔차에 반영되어 모형의 정확성을 저해했을 가능성이 있다.

둘째, 자기상관 문제를 완화하기 위해 적용한 1차 차분 기법은 시간적 종속성을 줄이는 데에는 효과적이었으나, 동시에 변수의 절대적 수준 정보를 간과한다는 한계를 가진다. 이에 따라 많은 설명변수의 통계적 유의성이 감소하였고, 고용률의 구조적 결정요인을 파악하는데 어려움이 발생하였다.

셋째, 정규성 가정이 완전히 충족되지 못한 점도 중요한 한계이다. 차분 이후에 Q-Q plot에서 정규성의 위반이 확인되었다. 이는 회귀계수의 추론의 정확도에 영향을 미칠 수 있으며 대체 기법의 적용을 고려할 필요가 있다.

마지막으로, 본 연구에서 사용한 변수들은 주로 국제기구에서 제공하는 지표 위주로 구성되어 있다. 이들 변수는 국가 간의 구조적 차이를 반영하기에는 유용하지만, 고용률에 직접 영향을 미치는 노동시장 정책, 사회보장제도, 산업별 고용구조, 기술 수용도 등 보다 미세한 정책·제도적 변수를 포함하지 못했다. 또한 일부 변수는 결측치를 중앙값으로 대치하여 사용했는데, 이는 데이터의 변동성을 다소 단순화하고 설명력에 영향을 미쳤을 수 있다.

8. 참고문헌

- [1] Solow, R. M. (1956). "A Contribution to the Theory of Economic Growth." *The Quarterly Journal of Economics*, 70(1), 65-94.
- [2] Lewis, W. A. (1954). "Economic Development with Unlimited Supplies of Labour." *The Manchester School*, 22(2), 139-191.
- [3] Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.

[4] Romer, P. M. (1990). "Endogenous Technological Change." *Journal of Political Economy*, 98(5), S71–S102.

데이터 출처

OECD Data Explorer, Retrieved from <https://data-explorer.oecd.org/>

World Bank Open Data, Retrieved from <https://data.worldbank.org/>

UNDP Human Development Reports, Documentation and downloads, Retrieved from
<https://hdr.undp.org/data-center/documentation-and-downloads>