

## 데이터 시각화 기법을 활용한 행복취약계층 찾기

구교준\*

김동욱\*\*

우리나라의 경제적 수준은 상위권인 반면 행복 수준은 중위권에 머물고 있다. 이러한 경제적 수준과 행복 수준의 괴리를 살펴보기 위해 본 연구에서는 행복 수준이 상대적으로 낮은 행복취약계층의 특성에 주목하고자 하였다. 데이터 시각화는 정보를 산점도·그래프 등과 같은 시각적인 요소로 나타내는 방법으로, 데이터 내 패턴 및 추세를 쉽게 파악할 수 있게 한다는 장점을 갖고 있다. 본 연구에서는 이러한 데이터 시각화 기법으로 선형 모형의 정보 손실을 보완하고, 보다 많은 정보에 기반하여 행복취약계층을 파악하고자 한다. 이를 위하여 통계청 사회조사와 행정연구원의 사회통합실태조사 자료를 통합하였다. 통합된 데이터는 행복 수준에 중요하게 작용하는 사회인구학적 요인인 성별, 연령, 교육, 건강, 소득, 취업여부, 혼인상태에 따라 3,360개의 소집단으로 나누고, 데이터 시각화 기법을 통해 행복취약계층을 시각적으로 제시하였다. 마지막으로 데이터 시각화 결과를 바탕으로 연령 및 성별에 따라 서로 달리 나타나는 니즈를 충족할 수 있는 정책이 수립되어야 함을 논의하였다.

주제어: 행복, 행복취약계층, 데이터시각화

\* 주저자, 교신저자. Univ of North Carolina at Chapel Hill에서 도시계획학 박사학위를 취득하고 현재 고려대학교 행정학과 교수로 재직 중이다. 주요 관심분야는 행복, 삶의 질, 지역 혁신, 창업 등이다(jkoo@korea.ac.kr).

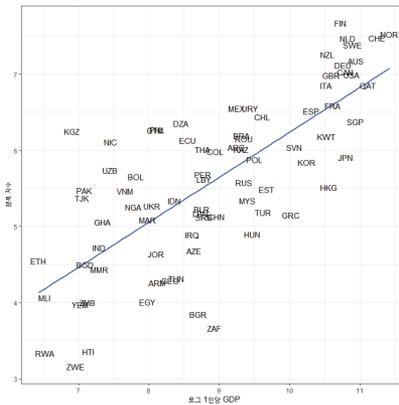
\*\* 공동저자. 현재 고려대학교 행정학과 석사과정에 재학중이며, 주요 관심분야는 행복, 저출생, 계량행정 등이다(dongwook1995@korea.ac.kr).

## I. 서론

유엔이 매년 발간하는 세계 행복보고서에 의하면 2021년 우리나라의 행복지수는 전 세계 146개국 중 59위이다(Helliwell et al., 2022). 중위권에 머무르고 있는 행복 수준과는 달리 우리나라의 경제적 순위는 2021년 GDP 기준 전 세계 10위이고 1인당 GDP 기준 전 세계 26위로 상당히 상위권에 속한다.

사실 경제적 수준과 행복 수준 간의 괴리는 이미 1970년대 Easterlin(1974) 이후 학계에서 많이 다루어졌던 문제이다. 미국 뿐 아니라 유럽의 주요 선진국과 일본에서 과거 수십년 동안 소득은 두 배, 세 배 증가하였지만 삶이 행복하다고 느끼는 사람들의 비율은 거의 변하지 않았음이 여러 연구를 통해 보고된 바 있다(Layard, 2005). 하지만 이러한 괴리는 다른 나라들과 비교했을 때 우리나라가 상당히 큰 편이고(구교준·임재영·최슬기, 2017), 이는 [그림 1]의 간단한 회귀분석에서도 잘 나타난다. 전 세계 주요국을 대상으로 1인당 GDP와 행복 간의 관계를 보여주는 [그림 1]에서 소득 수준에 따라 예상되는 행복 수준을 의미하는 회귀선에 비해 우리나라의 실제 행복 수준은 현저히 낮은 그룹에 속한다. 정반대의 경우라고 볼 수 있는 핀란드, 노르웨이, 네델란드 등 회귀선 위쪽의 국가들과는 큰 차이를 보이고 있다. [그림 1]이 보여주는 메시지는 명확하다. 소득이 행복을 결정짓는 가장 중요한 변수 중의 하나라고 할 때 우리나라에서 1달러의 소득이 가져다 주는 행복의 크기가 서구 선진국들에 비해 현저

〈그림 1〉 1인당 GDP와 행복수준 분포



출처: World Value Survey, 저자

히 작다는 점이다. 즉 행복을 가져오는 소득의 한계효율성, 혹은 달리 표현하면 소득의 질이 낮은 것이다.

국민이 행복한 나라가 좋은 나라라고 한다면, 이러한 분석은 뚜렷한 정책의 방향성을 제시해 준다. 우리나라가 소득에 비해 행복 수준이 현저히 낮은 점을 고려한다면 어떻게 1인당 GDP 3만 달러를 4만 달러로 높일 수 있을까만을 고민하는 것이 아니라 어떻게 1인당 GDP 3만 달러에서도 국민의 전반적인 행복 수준을 지금보다 높일 수 있을까를 좀 더 치열하게 고민해야 한다. 같은 소득이더라도 국민이 더 행복한 사회, 즉 소득의 양이 아닌 소득의 질을 지향하는 사회로의 방향 전환이 필요한 시점이다.

같은 소득이더라도 상대적으로 높은 행복 수준을 달성하기 위해서는 소득 이외에도 행복 수준 결정에 중요한 영향을 끼치는 경제, 정치, 사회적 요인들의 질적 향상을 위한 노력이 필요하다. 행복 결정요인에 대한 기존 연구는 이를 위해 어떤 정책적 방향성이 필요한지에 대한 중요한 함의를 제공해 준다. 경제적으로는 소득불평등을 줄이고(Bjornskov, Dreher, Fischer, & Schnellenbach, 2009; Ferrer-i-Carbonell & Ramos, 2014; Zagorski, Evans, Kelly, & Piotrowska, 2014; 장승진, 2011), 정치적으로는 민주주의와 자유를 더욱 공고히 하며(Inglehart, Foa, Peterson, & Welzel, 2008; Verme, 2009; 유정민·최영준, 2020), 사회적으로는 건강과 교육 수준의 향상을 위해 투자하고(Hartog & Hessel, 1998; Mustonen, 2009) 투명하고 서로 신뢰하는 환경을 만드는 것(Rothstein, 2010; Tokuda, Fujii, & Inoguchi, 2010) 등을 예로 들 수 있다.

다른 한편으로는 행복 수준이 높은 집단과 낮은 집단을 구분하여 낮은 집단의 삶의 질 향상에 정책의 초점을 맞추는 접근을 생각해 볼 수 있다. 행복 수준이 높은 집단은 이미 소득이 상대적으로 높고 건강이나 교육 등 행복 수준에 중요한 영향을 미치는 다른 환경 요인도 상대적으로 나은 편이다. 따라서 이들 요인의 질적 향상으로 인해 기대할 수 있는 행복 수준 향상의 여지는 그리 크지 않다고 할 수 있다. 게다가 행복이 경제학의 효용 개념의 연장선상에 있음을 고려하면, 한계효용(행복)의 차이로 인해 행복 결정요인이 행복 수준이 높은 집단과 낮은 집단에 다르게 작용할 수 있을 것이라는 점도 쉽게 예상할 수 있다. 그러나 아쉽게도 기존 연구에서 행복 수준이 높은 집단과 낮은 집단을 구분하여 분석한 경우는 거의 찾아볼 수 없다. 이를 위해선 먼저 행복 수준이 낮은 행복취약계층을 정의해야 하는데, 행복취약계층에 초점을 맞추어 이들이 어떤 사람들이고 어떤 특성을 가지고 있는지 찾으려는 시도는 지금까지 거의 없었다. 본

연구에서는 이와 같은 기존 연구의 한계점에 주목하여 행복취약계층을 찾는 분석을 시도하였다.

이러한 분석은 행복의 결정요인을 출발점으로 하는데, 행복 결정요인에 대한 대부분의 기존 연구들은 회귀분석이나 ANOVA 등의 통계적 모형을 기반으로 하고 있다. 통계 모형을 기초로 한 분석은 기본적으로 선형 효과를 가정하고 변수의 수를 최소화하며 설명력을 최대화하는 간결성의 원칙(principle of parsimony)을 기반으로 하고 있다. 따라서, 두 개 이상 요인들의 상호작용 또는 비선형적 관계를 살펴보기 어렵다. 더 나아가, 데이터가 존재하지 없음에도 불구하고 종속변수의 값을 추정할 수 있다는 점 또한 한계로 작용할 수 있다. 이는 통계적 방법론의 장점이긴 하나, 변수들의 상호작용 또는 비선형적 관계가 존재할 수 있고 실험실이 아니기 때문에 발생하는 선택편향(selection bias) 등과 같은 문제로 인해 추정에 편향과 비효율성이 발생할 수도 있다는 점에서 한계로도 지적된다.

이와 같은 기존 연구의 방법론적 한계점에 착안하여 본 연구에서는 행복취약계층을 찾는 분석을 위해 통계적 방법론이 아닌 데이터 시각화 기법을 활용한다. 데이터 시각화는 정보를 산점도, 그래프 등과 같은 시각적인 요소로 나타내는 방법이며, 데이터 내 패턴 및 추세를 쉽게 파악할 수 있게 한다는 장점이 있다. 데이터 시각화를 통해 선형모형의 간결성으로 인한 정보 손실을 보완하여 보다 많은 정보를 기반으로 행복 소외계층을 파악하는 것이 가능해진다. 이를 위하여 우리나라에서 행복 연구에 가장 폭넓게 활용되고 있는 통계청 사회조사와 행정연구원의 사회통합실태조사 자료를 통합하였다. 통합된 자료는 행복 수준에 중요하게 작용하는 사회인구학적 요인인 성별, 연령, 교육, 건강, 소득, 취업여부, 혼인상태에 따라 3,360 개의 소집단으로 나누고, 데이터 시각화 기법을 통해 행복취약계층을 시각적으로 제시하였다. 마지막으로 데이터 시각화 결과를 바탕으로 정책적 함의를 논의하였다.

## II. 이론적 논의

### 1. 행복의 정의와 측정

행복 연구의 가장 어려운 부분은 행복을 어떻게 정의하고 측정할 것이냐의 문제이다. 특히 행복의 정의 부분은 삶의 질, 효용, 웰빙 등 유사 개념들이 많고 '행복'이라

는 개념만 보더라도 그 안에 행운, 쾌락, 의미 등의 다양한 측면이 포함되어 있어 더욱 어려운 문제이다.

Oishi et al.(2013)은 그리스 시대에서부터 중세에 이르는 오랜 기간 동안 행복 개념의 중심에는 행운이 있었음을 설명하였다. 이는 중세 시대까지의 행복 개념은 현대의 행복 개념과는 많이 달랐음을 시사한다. 현대적인 의미의 행복 개념이 사회과학 담론에 처음 등장한 것은 Diener(1984, 1994)의 주관적 안녕(subjective well-being) 연구를 통해서이다. Diener의 주관적 안녕 개념은 행복을 인지적인 부분과 정서적인 부분으로 나누어 개념화하고 조작화함으로써 행복의 과학적 연구에 초석을 닦았다. 주관적 안녕 개념 중 특히 정서적 부분에 해당하는 개인의 긍정적인 감정은 그리스 철학의 hedonism 과 맞닿아 있다.

이와 달리 행복을 쾌락이 아닌 의미라는 측면에서 이해하는 접근도 있다. Sen(1980)과 Nussbaum(2003)은 삶의 역량(capability)을 개인이 의미와 가치를 두는 무언가를 자유롭게 추구하는 과정을 통해 느끼는 충만감이라고 정의하고 이를 통해 의미 중심의 행복 개념을 소개하였다. 행복을 쾌락이 아닌 의미 중심으로 이해하는 이러한 접근은 그리스 철학의 eudaimonism 에 가깝운데, 이는 주관적 안녕 개념의 인지적 부분인 삶의 만족(life satisfaction)이 어느 정도 반영하고 있다. 개인이 삶의 만족을 느끼는 메커니즘 안에는 자기 자신의 삶을 얼마나 자유롭게 선택하느냐가 가장 중요한 요소 중의 하나로 포함되기 때문이다. 정리하면, 행복의 정의에 대한 기존의 이론적 논의는 hedonism 에 바탕을 둔 쾌락으로서의 행복 중심의 개념과 eudaimonism 에 바탕을 둔 의미로서의 행복 중심의 개념으로 구분할 수 있는데, 주관적 안녕 개념은 hedonism 에 해당하는 정서적 평가와 eudaimonism 에 해당하는 인지적 평가를 모두 담고 있는 복합적인 개념이라고 할 수 있다(Landes, 2015).<sup>1)</sup>

행복의 측정은 해외의 경우 World Value Survey, Gallup World Poll, 그리고 국내의 경우 사회통합실태조사, 한국사회종합조사와 같은 대규모 설문조사를 통해 이루어진다. 이들 조사는 하루 동안 느낀 행복감의 정도나 지난 삶을 돌이켜 보았을 때 삶에 대해 만족하는 정도 등의 문항을 통해 행복을 측정한다. 다만 인지적 측면에 해당하는 삶의 만족에 대한 질문은 조사 유형과 관계 없이 어느 정도 표준화되어 유사한데, 정서적 측면에 해당하는 행복감 등 감정에 대한 질문은 조사마다 내용과 척도가 조금씩 상이하여 비교가 어렵다. 다만 Kahneman & Krueger(2006)는 정서적인 행복감과 삶의 만족이 강한 상관관계를 보인다고 설명한 바 있는데, 이는 많은 기존 연

1) 행복의 정의와 측정에 대한 보다 자세한 논의는 구교준 외(2015) 참조.

구들에서 행복 측정이 삶의 만족을 중심으로 이루어지는데 대한 이유와 정당성을 제공해 준다고 볼 수 있다. 본 연구의 행복 측정도 사회통합실태조사와 한국사회종합조사의 삶의 만족에 대한 측정 문항을 활용하였다.

## 2. 행복과 사회인구학적 요인

본 연구에서는 행복취약계층을 찾기 위하여 행복 수준에 중요하게 영향을 끼치는 것으로 알려진 성별, 연령, 교육, 건강, 소득, 취업여부, 혼인상태의 7 가지 사회인구학적 요인을 활용하였다. 이들 변수는 많은 행복 연구에서 다루어져 왔으며, 연구마다 서로 다른 결과와 해석을 내놓고 있다. 같은 변수라도 연구자와 자료에 따라 통계적으로 유의하게 행복 수준에 긍정적인 영향을 미치기도 하고 부정적인 영향을 미치기도 하는 것으로 나타났다. 또한 비선형적 관계 또는 통계적으로 유의하지 않은 결과가 보고되기도 하였다. 변수 간 상호작용이 문제라고 지적하는 연구가 있는 반면 그러한 관계를 부정하는 연구 또한 존재하는 등 그 중요성에도 불구하고 뚜렷한 패턴을 발견하기는 쉽지 않다.

**성별:** 성별은 행복과 다양한 관계를 갖고 있다. 행복을 인식하는 인지체계가 성별에 따라 달라질 수도 있으며, 사회경제적 지위 및 자원에 대한 접근과 활용 정도 등과 같은 요소들 또한 성별에 의해 차이가 발생할 수 있기 때문이다. 많은 기존 연구들이 다양한 자료를 활용하여 여성의 행복 수준이 남성에 비해 일반적으로 높다는 점을 보여 주었다(Graham & Chattopadhyay, 2013; Helliwell, Layard, & Sachs, 2019; Zweig, 2015). 하지만 남녀 간에 존재하는 응답의 기준점이 상이할 수 있음을 고려하고 이를 정규화한 연구에서는 그 반대의 결과도 보고된 바 있다.

**연령:** 상식적으로 건강상태, 소득수준, 교육수준, 혼인상태와 같은 요인들은 연령과 밀접한 관계를 연령과 밀접한 관계를 가지고 있기 때문에, 연령이 행복과도 높은 상관 관계를 가질 것이라고 쉽게 예상할 수 있다. 기존 연구에서는 일반적으로 연령과 행복이 U자형 관계를 가진다고 보고되고 있다(Blanchflower & Oswald, 2004; Clark & Oswald, 1994). 즉 청소년기에 높은 행복 수준을 보이고 중장년기를 거치면서 행복 수준의 하락을 경험하다가 노년기에 다시 행복 수준이 상승한다는 것이다.

**교육:** 교육은 삶의 여러 영역 중 행복에 가장 중요한 영향을 줄 수 있다. 교육 수준이 높아짐에 따라 직업을 선택하는 폭이 넓어지거나, 더욱 높은 소득과 안정적인 일자리를 얻을 수 있다. 자유로운 직업의 선택과 안정된 일자리는 행복의 중요한 필요조건이다. 또한 이와 같이 경제적인 측면 뿐 아니라, 교육을 통해 개인은 삶을 평가할 때

보다 다양하고 넓은 관점을 가질 수 있게 된다. Frey & Stutzer(2000)는 교육 수준에 따라 행복 수준이 달라질 수 있음을 잘 보여주었다. 특히 Helliwell & Putnam(2004)은 높은 교육 수준이 건강한 삶이나 고소득으로 이어져 행복에 긍정적인 영향을 미친다고 설명하고 있다.

건강: 건강은 행복의 가장 중요한 필요조건 중 하나라고 할 수 있다. 일을 하거나 다른 사람들과 교류하기 위해서 건강이 뒷받침되지 않으면 근본적으로 어렵기 때문이다. 기존 연구에서는 신체적 장애와 같은 건강 수준의 변화는 완벽한 적응(adaptation)의 대상이라고 여기지기도 했지만, 최근에는 객관적 주관적 건강 수준에 따라 행복 수준이 변화하며, 따라서 건강의 변화는 완전히 적응되지 않는다는 연구 결과가 보고되고 있다(Easterlin, 2003).

소득: 소득과 행복 간의 관계는 행복 연구에서 가장 많은 분석이 이루어진 주제이다. 경제학적으로 본다면 보다 많은 소득은 보다 많은 소비로 이어지고 이는 자연스럽게 보다 높은 수준의 효용, 즉 행복으로 연결된다. 실제로 수많은 실증 연구들이 적어도 일정 수준까지는 소득 증가가 행복 수준 향상으로 이어진다는 사실을 보여주었다(Diener, Sandvik, Seidlitz, & Diener, 1993; Easterlin, 2001). 다만 이러한 경향성은 소득이 증가함에 따라 점차 약해지는 것으로 나타났다(Lane, 2000).

취업여부: 취업여부는 소득과 같은 경제적이고 물질적인 요소 뿐 아니라 자아실현과 같은 정신적이고 비물질적인 요소와도 깊은 관련이 있다. 따라서 취업은 일자리와 차별이 그 이상의 의미를 가진다. 이러한 이유로 기존의 많은 연구들이 실업이 행복에 매우 부정적인 영향을 미칠 수 있음을 확인하였다(Blanchflower & Oswald, 2004; Clark & Oswald, 1994). 특히 Clark & Oswald(1994)의 연구는 실업이 이혼이나 별거와 같은 요인보다 정신 건강에 더 부정적인 영향을 미칠 수 있음을 확인하였다.

혼인상태: 많은 행복 연구에서 배우자 여부를 행복에 크게 영향을 미치는 요인으로 보고 있다. 결혼을 통해 가구 소득의 상승, 안정감, 자아실현 등과 같은 경제적·비경제적인 혜택을 얻을 수 있기 때문이다. 결혼은 누군가 강제하는 것이 아니라 개인이 스스로 결정한다는 점에서 Becker(1975)는 결혼을 선택한 사람들의 효용이 미혼으로 남은 사람들의 효용보다 높을 것으로 해석하기도 하였다. Easterlin(2003)은 코호트내 기혼자와 미혼자 사이에 본질적인 차이가 없으며 기혼자의 행복 수준이 미혼자의 행복 수준보다 높게 나타난 것을 두고 결혼에는 완벽한 적응이 일어나지는 않는다고 설명하였다. 또한 Helliwell & Putnam(2004)은 행복과 사회적 자본의 연결고리를 살펴보았는데, 혼인 상태로 측정된 가정 수준의 사회적 자본은 행복과 강력한 관계를 갖는 것으로 나타났다.

### Ⅲ. 데이터와 분석방법

#### 1. 분석 데이터 소개

우리나라의 행복 수준을 다루고 있는 데이터는 여러 가지 있다. 설문조사 데이터로는 해외의 World Value Survey(WVS), 국내의 한국노동패널조사(KLIPS, 한국노동연구원), 사회조사(통계청), 사회통합실태조사(한국행정연구원), 한국종합사회조사(KGSS, 성균관대 서베이리서치 센터) 등이 있다. 본 연구에서 사용할 데이터는 통계청의 사회조사와 한국행정연구원의 사회통합실태조사 데이터이다. 이 두 데이터는 1) 주관적 웰빙에 대한 설문을 포함하고, 2) 매년 조사가 시행되며, 3) 패널 데이터가 아니라는 공통된 특징을 가지고 있다.

먼저 통계청 사회조사는 1977년부터 매년 5월 중순에 실시하고 있는 설문조사로, 국민의 삶의 질과 관련된 총 10개 영역에 대해 2년 주기로 조사가 진행된다. 매년 포함되는 기본적인 항목에 더하여 홀수 연도에는 복지·사회참여·여가·소득과 소비·노동에 대한 사항을, 짝수 연도에는 가족·교육과 훈련·건강·범죄, 안전·생활환경에 대한 사항을 모듈로 추가하여 조사하고 있다. 또한 표본추출틀로 전국 모든 인구의 정보를 담고 있는 통계청 등록센서스를 사용하고 표본추출 방법으로는 층화 이단계 집락추출법을 활용하여 전국민에 대한 높은 대표성을 확보하고 있다. 2021년 통계청 사회조사의 총 표본가구수는 18,576 가구이며, 총 표본수는 36,423 명이다. 사회조사는 2009년부터 매년 주관적 만족감 문항을 포함하여 조사를 실시하고 있는데, 2021년 설문에서 주관적 만족감을 묻는 문항은 “귀하는 자신의 생활을 전반적으로 고려할 때, 현재 삶에 어느 정도 만족하십니까?”로 제시되어 있다(5점 리커트 척도).

사회통합실태조사는 한국행정연구원에서 2011년부터 매년 9월에 실시하고 있는 설문조사로, 사회의 통합 수준에 대한 국민들의 인식 및 태도를 파악하는 것을 목적으로 하고 있다. 사회통합실태조사는 주관적 웰빙 이외에도 사회참여, 정치참여, 사회적 소통, 신뢰, 거버넌스, 공정성, 관용성, 사회보장 등의 조사항목을 포함하고 있다. 사회통합실태조사의 표본추출틀과 표본추출 방법은 통계청 사회조사와 동일하게 통계청 등록센서스와 층화 이단계 집락추출법을 활용하고 있으며, 총 표본가구 수는 약 4,000 가구, 총 표본수는 약 8,000명 정도이다. 사회통합실태조사에서는 2013년부터 행복과 주관적 웰빙에 대한 설문문항을 포함하였는데, 2021년 문항을 기준으로 각각 “귀하는 어제 어느 정도 행복했다고 생각하십니까?”와 “귀하는 요즘 자신의 삶에 대해 전반적으로 얼마나 만족하고 있습니까?”라고 제시되어 있다(0-10 척도).

본 연구는 약 3,000 여개의 소집단을 나누어 행복을 분석하는 관계로 최대한 많은 수의 사례 확보가 매우 중요하다. 따라서 앞서 소개한 통계청 사회조사와 행정연구원 사회통합실태조사의 최근 5년 데이터를 통합하여 분석에 사용하였다. 행복의 측정을 위해 두 조사 모두 삶의 만족도에 대한 질문을 포함하고 있으며, 구체적인 질문의 내용은 통합이 가능할 정도로 유사한 수준이라고 판단하였다. 또한 패널 데이터가 아니라 매년 증복되지 않은 샘플을 대상으로 조사가 이루어지기 때문에, 5년치 데이터를 통합함으로써 커버가 가능한 소집단의 수를 극대화 시킬 수 있었다.

한편 사회조사와 사회통합실태조사에 포함된 인구사회학적 변수들은 다음과 같이 재코딩하여 분석에 사용하였다. 우선 연령은 20 대, 30 대, 40 대, 50 대, 60 대 이상으로 구분하였다. 교육 수준은 고졸 이하, 2 년제대학 졸, 4 년제대학 졸, 대학원 이상으로 구분하였다. 건강은 주관적 건강에 대한 응답을 기초로 건강함과 건강하지 않음으로 구분하였고, 소득은 월소득 100 만원 이하에서 600 만원 이상까지 100 만원 단위로 구분하였다. 취업여부는 정규직 비정규직 구분 없이 취업과 미취업으로 구분하였고, 혼인상태는 미혼, 기혼, 이혼 및 사별로 구분하였다. 데이터를 재코딩한 후 성별, 연령, 교육, 건강, 소득, 취업여부, 혼인상태에 따라 3,360 개의 소집단으로 나눈 후 해당 집단에 속한 개인들의 삶의 만족도의 평균을 구하였다. 이 때 해당 집단에 포함된 인원이 10 명 미만일 경우 분석의 정확성을 위해 분석에서 제외하였다.

## 2. 통계적 모형과 데이터 시각화

대부분의 행복 연구들은 기본적으로 회귀모형과 ANOVA 등과 같은 선형 모형을 기반으로 분석이 이루어졌다. 이러한 통계적 모형들은 최소한의 변수로 최대한의 분산을 설명한다는 간결성(parsimony)의 원칙을 기반으로 하고 있는데, 간결성의 원칙은 “보다 간단한 이론이 보다 높은 사전 확률을 갖고 있다.”라는 원칙을 의미하며 (Jeffreys, 1998), 세상을 이루는 구조와 인과관계 및 과정이 간단하다는 가정을 의미한다(King, Keohane, & Verba, 1994). 이처럼 간결성의 원칙은 복잡하게 얽힌 사회현상에서 핵심적인 원인들이 관심이 되는 현상에 미치는 영향을 살펴볼 수 있도록 해준다는 점에서 연구를 용이하게 한다는 장점이 있다. 그러나 이러한 통계적 선형모형에는 간결성의 원칙이라는 장점과 함께 통계 모형 그 자체로 인한 한계점도 존재한다. 우선, 간결성의 원칙으로 인해 변수들의 상호작용을 충분히 고려하기 어렵다는 점이 있다. 간결성의 원칙은 최소 변수로 최대 변수를 설명한다는 뜻도 있으나, 모형의 해석이 쉬워야 한다는 뜻도 내포하고 있기 때문이다. 따라서 회귀분석에서 사용되는

상호작용항은 2 차항이 최대로 3 차항 이상은 잘 사용되지 않는 편이며, 로지스틱 회귀 분석의 경우 상호작용에 대한 해석이 난해해지기 때문에 매우 드물게 사용된다. 이처럼 간결성의 원칙은 현실을 단순화함에 따라 나타나는 정보의 손실을 감수하면서 모형의 일반성과 해석의 용이성을 동시에 추구하는 원칙이라고 할 수 있다.

여기에 더해 선형모형 그 자체에서 오는 한계점도 존재한다. 이는 크게 두 가지로 나눌 수 있는데, 하나는 선형모형의 가정에서 유발되는 한계점이며 다른 하나는 모형이 평균 또는 기대값을 기반으로 하고 있다는 점에서 발생하는 한계점이다. 전자는 흔히 사용되는 회귀분석(최소자승법)에서 독립변수 간의 통계적 독립성과 종속변수와 독립변수 간의 선형관계 가정을 의미한다. 앞선 행복 연구들을 살펴보면 행복 영향요인들이 선형적일 뿐 아니라 비선형적인 경우도 흔하게 존재한다. 변수들이 밀접한 관계를 갖고 있는 경우에는 주성분회귀(principal component regression)를 진행할 수 있으나, 주성분회귀 역시 변수들 간의 선형적 관계를 가정하고 있기 때문에 비선형적 관계의 경우 문제가 발생할 수 있다. 또한 독립변수와 종속변수 간에 선형관계 가정은 회귀분석의 핵심적인 가정인 다변량정규분포 가정 때문에 필요한데, 비선형적 관계가 존재하는 경우 회귀식 추정시 비효율성의 문제가 발생한다. 이러한 문제를 해결할 수 있는 방안으로는 generalized least square(GLS) 또는 비모수적 추정 방법을 적용할 수 있으나, 모형의 해석이 난해해진다는 점에서 간결성의 원칙과 충돌할 수 있다.

한편 선형모형은 평균 또는 기대값을 기반으로 회귀식을 추정하기 때문에, 데이터가 부족하더라도 효율적인 추정을 할 수 있다는 장점이 있다. 실험계획에서 사용되는 일부실시법(fractional factorial design)<sup>2)</sup> 같은 기법이 그 예시가 될 수 있는데, 이 또한 각 변인으로 인한 효과가 선형적임을 가정하고 있어 상호작용과 같은 비선형적 관계를 고려하기 어려울 수 있다. 더 나아가, 이러한 추정 방식은 연구자가 임의로 처치와 통제변수를 설정할 수 있는 실험에서는 효율적일 수 있으나 이러한 처리가 불가능한 사회 과학의 경우에는 선택편향(selection bias)의 가능성이 존재한다. 따라서, 변

2) 일부실시법은 주된 관심 사항이 되는 처리(treatment)가 많을 때, 실험 횟수를 줄여 보다 효율적인 실험을 설계하고자 할 때 사용되는 방법이다. 자주 사용되는  $2^k$  실험계획법을 예시로 들 경우, 처리의 개수가  $k$ 개일 때 고차의 상호작용까지 고려하는 완전한 실험을 하고자 할 경우  $2^k$  번의 실험을 해야 하며 처리의 수가 증가할수록 실험을 해야 하는 조합의 수 또한 기하급수적으로 증가함을 알 수 있다. 만약 주된 관심이 있는 처리의 수가 10개일 경우, 총 실험의 수는  $2^{10} = 1,024$  번이 될 것이다. 그러나 일부실시법은 고차의 상호작용을 처리의 주효과와 보다 낮은 차수의 상호작용에 교란(confounding)시키는 대신 실험의 조합 수 또는 실험 횟수를 줄일 수 있다. 해당 방법론에 대한 보다 자세한 설명은 박성현, 2012. <<현대실험계획법>>. 서울: 민영사 11장을 참고.

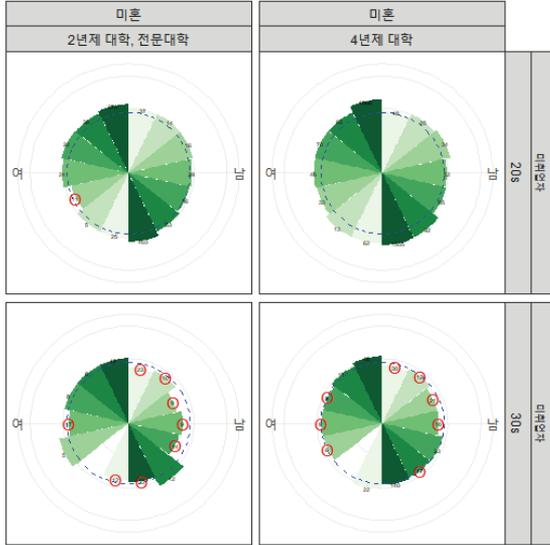
수들이 비선형적 관계를 갖거나 데이터에 편향이 존재하는 경우 기댓값과 같은 방식에 의존하는 통계적 모형에 타당성 문제가 제기될 수도 있다.

이처럼 흔히 사용되는 통계적 모형은 해석이 용이하고 강력한 설명력을 가진다는 장점과는 별개로 간결성의 원칙으로 인한 정보 손실과 모형의 가정에서 발생하는 한계를 지니고 있으며, 이로 인해 전체 그림을 보지 못하는 문제와 잘못된 추정을 할 수 있다는 문제가 발생할 수 있다. 이러한 이유로 인해 본 연구에서는 간결성의 원칙과 선형 모형이 아닌 데이터 시각화를 분석 방법으로 채택하였다. 데이터 시각화는 데이터가 담고 있는 정보를 시각적인 요소로 변환하여 제시하는 방법으로, 일반적인 회귀 분석에 비해 정보 손실이 적고 시각화를 통해 패턴 및 추세를 쉽게 파악할 수 있다.

#### IV. 데이터 시각화 분석결과: 누가 행복취약계층인가?

앞 절에서 설명하였던 과정을 거친 데이터를 기반으로 데이터 시각화를 진행하였으며, 그 과정에서 가능한 많은 정보를 담고자 하였다. [그림 2]는 혼인여부, 교육, 성별, 소득 등의 다양한 변수들이 어떻게 시각화 과정을 통해 행복이라는 종속변수와 연결되어 해석 가능한지 잘 보여준다. 그림에서 횡단은 혼인여부와 교육을 기준으로, 종단은 취업여부와 연령을 기준으로 그래프를 나누었다. 각 패널 안에 있는 나이팅게일 로즈 다이어그램(Nightingale rose diagram)은 성별, 가구소득 및 행복 수준에 대한 정보를 담고 있다. 성별은 그래프의 좌우(우 = 남성, 좌 = 여성)로 표현되었고 가구소득은 색이 진할수록 소득이 높은 가구 소득을 의미한다. 또한 부채꼴의 길이는 본 연구의 주된 관심사인 행복 수준을 뜻한다. 따라서 부채꼴이 밖으로 뻗을수록 해당 집단의 행복 수준이 높다는 것을 의미한다. 한편 각 소집단에 속하는 표본 수를 부채꼴의 바깥 부분에 숫자로 기재하였다. 마지막으로 행복 소외계층을 시각화하기 위하여 행복 수준 하위 25%를 점선으로 표기하고, 여기에 속하는 집단을 빨간색 원으로 표시하였다. 즉 빨간색 원이 부채꼴 끝의 점선에 표시된 경우, 해당 집단이 하위 25% 또는 행복 소외계층에 속한다는 것을 의미한다.

〈그림 2〉 데이터 시각화 예시



[그림 3]은 2017 년에서 2021 년까지 최근 5 년 간의 사회조사와 사회통합실태조사 데이터를 통합하여 데이터 시각화 기법을 통해 인구사회학적 변수를 기준으로 나누는 소집단별 행복 수준을 나타낸 결과이다. 여기에는 건강을 제외한 6 개 변수가 사용되었고, 통합 데이터의 총 260,211 개의 사례가 1,680 개의 소집단을 형성하였다. 이들 소집단은 해석의 편의를 위해 혼인여부와 취업여부를 기준으로 다시 6 개의 대그룹으로 구분하였다. 그래프의 해석은 편의를 위해 20, 30 대 청년층, 40, 50 대 장년층, 그리고 60 대 이상 노년층으로 구분하여 진행하였다.

### 1. 20, 30대 청년층

[그림 3]에서 가장 눈의 띄는 부분은 연령, 학력에 의한 행복 수준의 차이는 그다지 뚜렷하게 관찰되지 않지만 혼인상태와 취업여부의 영향이 두드러지게 나타난다는 점이다. 이러한 특성은 그룹 1(미혼\*실업/미취업)과 그룹 3(이혼/사별\*실업/미취업), 그룹 6(이혼/사별\*취업)에 행복취약계층이 상대적으로 집중되어 있다는 점에도 잘 나타난다. 그룹 1 을 다른 그룹과 비교해서 살펴보면 동일한 학력과 연령대에서 미혼이면서 실업/미취업인 집단 중에 행복취약계층이 눈에 띄게 많은 것을 알 수 있다. 그룹 3 에서도

비슷한 패턴이 관찰되는데, 동일한 학력과 동일한 연령대를 비교할 때 이혼/사별이면서 실업/미취업인 집단 중에 상대적으로 행복취약계층이 많은 것을 알 수 있다. 이 두 그룹에 해당 사례가 전혀 없는 집단이 다수임을 고려하면 행복취약계층이 얼마나 집중적으로 분포되어 있는지 알 수 있다. 특히 주목할 점은 가계소득이 꽤 높은 집단 중에도 행복취약계층으로 분류되는 사례가 상당 수 관찰된다는 점이다. 소득이 일정 수준 되더라도 혼인상태나 취업상태로 인해 발생하는 행복 수준의 저하를 피하기 쉽지 않음을 의미한다.

한편 그룹 1에서 2년제 대졸 이상 20대는 실업/미취업 상태로 인한 부정적인 영향을 그리 크게 받지 않는 것으로 보인다. 하지만 30대 이상에선 연령이나 성별에 관계 없이 실업/미취업 상태의 부정적인 영향이 상당히 두드러진다. 취업상태가 행복에 미치는 영향은 그룹 1을 그룹 2, 그룹 4와 비교함으로써 더욱 뚜렷하게 볼 수 있다. 먼저 그룹 1(미혼\*실업/미취업)과 그룹 4(미혼\*실업/취업)에서는 연령과 학력, 혼인상태를 동일하게 두고 취업여부의 효과를 비교할 수 있다. 특히 20, 30대에서 행복수준이 하위 25%이하에 속하는 행복취약계층의 차이가 두드러지는데, 그룹 1의 경우 소득 수준에 상관 없이 케이스의 수가 충분한 거의 모든 소집단이 행복취약계층으로 나타난 반면, 그룹 4에서는 동일한 조건의 행복취약계층이 현저히 줄어든 것을 볼 수 있다. 이는 20, 30대 청년층에서 취업여부가 행복에 얼마나 중요한 역할을 하고 있는지 잘 보여준다.

한편 청년층에서 두드러지게 관찰되는 실업/미취업 상태의 부정적인 웰빙 효과는 혼인 여부에 따라 다소 영향을 받는 것으로 보인다. 그룹 1(미혼\*실업/미취업)과 그룹 2(기혼\*실업/미취업)에서는 연령, 학력, 취업여부를 동일하게 두고 혼인상태의 효과를 비교할 수 있다. 두 그룹은 동일하게 실업/미취업상태임에도 불구하고 30대 미혼 집단과 기혼 집단을 비교하면, 기혼 집단에서의 행복취약계층이 미혼 집단에서 보다 현저히 적음을 알 수 있다. 결혼을 한 집단의 경제적 형편이 그렇지 않은 집단에 비해 상대적으로 나은 것이라는 선택적 편향의 가능성을 고려해 볼 수 있으나, 제시된 다이어그램이 가계소득까지 포함하여 통제된 결과임을 고려하면 선택적 편향의 가능성은 그리 높지 않다고 볼 수 있다.

## 2. 40, 50대 중장년층

결혼과 관련하여 한가지 흥미로운 사실은 30대에서 발견되는 결혼이 가진 실업/미취업의 부정적 영향에 대한 상쇄효과가 40, 50대 중장년층에서는 상당 부분 희석되는

것으로 보인다는 점이다. 그룹 1, 2, 3(미혼 vs. 기혼 vs. 이혼/사별로 구분되는 실업/미취업 집단)의 40, 50 대를 비교하면 거의 유사하게 결혼 여부와 상관 없이 소득수준이 낮은 집단에서 집중적으로 행복취약계층이 나타난다. 30 대가 대부분 혼인 초기임을 감안하면 이는 결혼의 긍정적인 효과가 40, 50 대에서 포화효과(saturation effects)에 직면하는 것으로 해석된다. 이러한 패턴은 그룹 4, 5, 6(미혼 vs. 기혼 vs. 이혼/사별로 구분되는 취업 집단)에서도 발견되는데, 40, 50 대 취업자 집단에서 혼인 상태와 상관 없이 저소득층을 중심으로 행복취약계층이 집중되어 있음을 알 수 있다. 이는 중년으로 접어들면서 행복 결정요인으로 결혼보다는 소득의 중요성이 상대적으로 증가함을 의미한다고 해석할 수 있다.

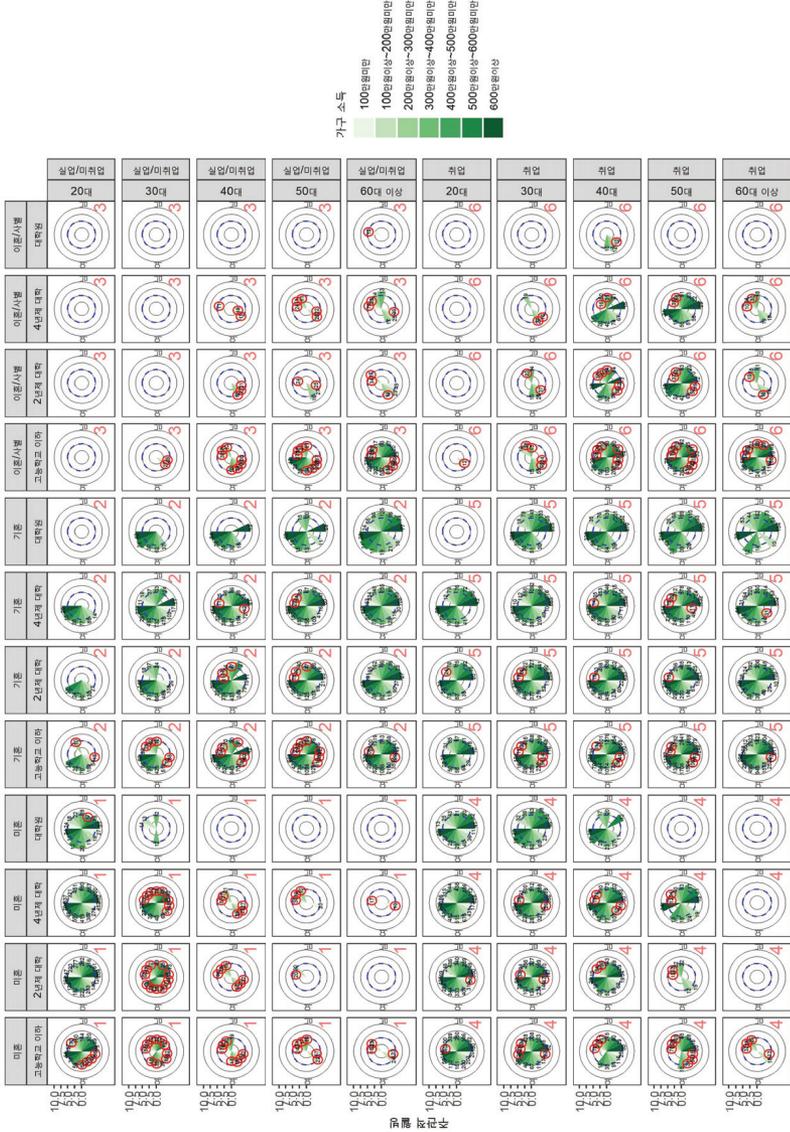
한편 40, 50 대 중장년 집단의 행복에서 결혼이 가지는 영향력은 줄어들지만 일자리와 소득의 영향력은 커진다는 사실도 관찰된다. 그룹 1, 2(미혼 vs. 기혼으로 구분되는 실업/미취업 집단)의 40, 50 비교와 그룹 4, 5(미혼 vs. 기혼으로 구분되는 취업 집단)의 40, 50 대 비교는 각각 중년의 미취업자와 취업자 집단에서 결혼여부가 큰 차이를 만들지 않는다는 점을 보여준다. 하지만 그룹 1, 2의 40, 50 대와 그룹 4, 5의 40, 50 대를 비교하면 취업자 그룹의 행복취약계층이 미취업자 그룹에 비해 상대적으로 적음을 알 수 있다. 특히 그룹 4(미혼\*취업)와 그룹 5(기혼\*취업)에서는 학력, 연령, 혼인여부의 차이가 그리 뚜렷하게 보이지 않는다. 즉 이는 일자리가 있고 어느 정도의 소득활동이 이루어지게 되면 다양한 인구사회학적 변수들이 행복에 미치는 영향이 상당 부분 줄어들게 됨을 의미한다. 하지만 그룹 6(이혼/사별\*취업)의 경우는 이야기가 좀 다르다. 그룹 4, 5와 비교하여 30 대 이상 전 연령대, 모든 교육수준에서 그룹 6에서 행복취약계층이 두드러지게 많이 관찰된다. 그 차이는 특히 40, 50 대 남성에서 좀 더 두드러지게 관찰된다. 40,50 대 여성의 경우엔 소득 최하위 계층에서만 행복취약계층에 나타나는 반면에 40, 50 대 남성의 경우엔 중간 소득 이상의 계층에서도 행복취약계층이 나타난다. 이는 중년 여성 집단에선 소득이, 그리고 중년 남성 집단에서는 이혼이나 사별을 통한 배우자 상실 여부가 행복에 중요하게 작용함을 의미한다.

### 3. 60대 이상 노년층

60 대 이상 노년층에서 관찰되는 가장 큰 특징은 성별, 학력, 소득, 취업여부 등 대부분의 인구사회학적 변수들이 행복취약계층의 분포에 그리 중요한 영향을 미치지 않는다는 점이다. 하지만 여러 가지 변수 중 혼인상태, 특히 이혼/사별의 영향력은 다른 연령대에 비해 두드러지게 관찰된다. 그룹 2, 5(기혼자 집단)와 그룹 3, 6(이혼/사별

〈그림 3〉 행복취약계층(혼인/교육/취업/연령 기준)

17년 ~ 21년 사회조사·사회통렬설문조사 데이터  
하위 25% 표기



• 동상권 우측 하단의 숫자는 본문에서 언급한 그룹 1-그룹 6 을 의미함

집단)을 비교하면 그 차이가 명확하게 드러난다. 기혼자 집단은 취업여부와 상관 없이 행복취약계층이 거의 나타나지 않는 반면에 이혼/사별 집단에서는 중간 소득 이하 집단에서 특히 많은 행복취약계층이 관찰된다. 특이한 점은 취업이 된 경우 60대 이상 여성은 이혼/사별 상태에서도 행복취약계층이 눈에 띄게 드러나지는 않지만 60대 이상 남성의 경우 취업이 된 상대라 하더라도 소득수준에 상관 없이 행복취약계층이 집중적으로 관찰된다는 사실이다. 이혼이나 사별로 인한 관계의 단절이 가져오는 부정적인 영향이 남성의 경우 훨씬 크며, 소득의 상쇄효과도 기대하기 어렵다는 점을 시사한다. 이는 60대 이상 노년층, 그 중에서도 특히 남성 이혼/사별 집단의 정신적 웰빙에 사회적으로 좀 더 많은 관심을 기울여야 함을 시사한다.

#### 4. 건강과 행복취약계층

지금까지 [그림 3]에서는 분석의 편의를 위해 횡단과 종단에 각각 두 개의 변수를 배치하여 집단을 구분하였다. [그림 4]에서는 [그림 3]의 집단 구분을 좀 더 세분화 시켜서 살펴보려 한다. 횡단 구분 기준에 건강을 추가하여 7개의 변수로 3,360개의 소집단과 12개의 대그룹을 생성하였다. 이러한 접근은 내용적으로 행복의 분포를 좀 더 세세하게 관찰할 수 있다는 장점과 함께, 방법론적으로는 기존 연구에서 중요한 행복 결정요인의 하나로 등장하는 건강의 효과를 통제한다는 의미 또한 가지고 있다. 즉 전체 집단을 건강하다고 응답한 집단과 건강하지 않다고 응답한 집단으로 구분하여 나이 탕제일 로즈 다이어그램을 그리면 건강의 영향을 제거한 상태에서 각각의 변수의 변화에 따라 행복취약계층의 분포가 어떻게 달라지는지 관찰하는 것이 가능하다.

우선 [그림 4]에서 드러난 가장 두드러진 특징은 행복취약계층이 건강이 좋지 않은 집단에 집중적으로 분포하여 발견된다는 점이다. 몇몇 예외를 제외하면 건강이 좋지 않다고 응답한 소집단은 연령이나 성별, 교육수준, 취업여부, 소득수준 등에 상관 없이 거의 대부분 행복 수준 하위 25%의 행복취약계층으로 분류되었다. 이는 건강이 개인의 행복 결정에 얼마나 중요한 요소로 작용하고 있는지 잘 보여준다. 건강은 이와 같이 다른 변수들의 웰빙 효과를 상쇄하고도 남음이 있다. 특히 기혼자 중 실업/미취업이나 저소득 집단에서 발견되던 행복취약계층은 건강을 구분하는 변수를 추가하여 집단을 좀 더 세부적으로 나누자 건강하다고 응답한 사람들 중에서는 대부분 사라졌다. 즉 결혼한 상태이고 건강하다면 거의 대부분의 사람이 행복취약계층으로부터 자유롭다는 것을 알 수 있다. 이에 반하여 건강하지 않다면 연령, 성별, 교육수준, 취업여부, 소득수준 등 어떤 변수에도 상관없이 행복취약계층에 속할 가능성이 크다.



한편 건강하지 않은 집단임에도 불구하고 행복취약계층에 속하지 않는 몇몇 예외적인 집단의 사례들은 다른 한편으로 행복에 있어서 소득이 가지는 중요성을 잘 보여준다. [그림 4]에서 건강하지 않음에도 행복취약계층에 속하지 않는 집단은 거의 월소득 500만원 이상의 고소득 집단임을 알 수 있다. 즉 이는 고소득층은 나쁜 건강의 부정적인 웰빙효과를 지속적인 관리를 통해 어느 정도 상쇄할 수 있음을 의미한다.

또 다른 주목할만한 패턴은 [그림 4]에서 건강하다고 응답한 사람들만 포함된 그룹 1의 30대 집단에 교육수준에 상관 없이 여전히 행복취약계층이 집중적으로 분포한다는 점이다. 그룹 1은 미혼\*실업/미취업 집단인데, 그룹 1의 30대는 그룹 2(기혼\*실업/미취업)와 그룹 4(미혼\*취업)의 30대와 비교할 때 확연히 구분된다. 그룹 2와 그룹 4의 30대에서는 그룹 1의 30대에서의 같은 행복취약계층의 집중적인 분포가 관찰되지 않는다. 즉 30대 청년 집단에서 집중적으로 관찰되는 행복취약계층은 결혼이나 취업을 하는 경우 상당 부분 해소되는 것으로 보인다. 이는 청년층의 행복 수준을 결정하는 가장 중요한 변수가 무엇인지 확연히 보여주는 결과라고 할 수 있다. 건강의 영향을 고려한 이후에도 청년층에게 있어 취업과 결혼의 영향은 매우 중요한 것으로 보인다.

한편 건강하지 않다고 대답한 응답자들을 분리하여 건강의 효과를 통제한 후에 결혼이 가지는 긍정적인 웰빙 효과가 더욱 뚜렷하게 관찰되었다. [그림 3]과 [그림 4]의 그룹 2(기혼\*실업/미취업)와 그룹 5(기혼\*취업)는 결혼한 응답자들의 행복 수준이 학력, 연령, 취업여부 등 다른 변수들에 따라 어떻게 달라지는지 보여준다. 건강의 효과를 통제하지 않은 [그림 3]에서는 기혼자 집단에서도 고졸 이하, 특히 실업/미취업 집단에서 행복취약계층이 상당히 나타난다. 그런데 이러한 패턴은 건강한 응답자만을 포함한 (그래서 건강의 효과를 통제) [그림 4]에서는 확연히 달라진다. 기혼자 집단에서는 취업여부나 소득에 상관없이 고졸 이하 집단에서도 행복취약계층이 거의 관찰되지 않는다. 이러한 사실은 [그림 3]의 40, 50대 중장년 집단에서 중요하게 등장한 일자리와 소득의 효과가 건강 변수의 부재로 인해 왜곡되어 있을 가능성을 시사한다. 건강하다고 응답한 집단만을 관찰하였을 때 일자리와 소득에 따른 행복취약계층 분포의 차이가 사라졌다는 사실이 그 가능성을 강하게 시사한다. 이는 관찰 가능한 상호작용에 한계가 있는 일반적인 회귀분석에서는 놓치기 쉬운 특징이다. 결론적으로 [그림 4]의 그룹 2와 그룹 5는 기혼자 집단에서 다른 변수와 관계없이 행복취약계층으로 분류되는 경우가 거의 없음을 잘 보여준다. 그리고 이는 그만큼 결혼이 가지는 긍정적인 웰빙효과가 상당한 수준임을 암시한다. 이러한 사실을 통해 또한 학력과 건강 간에 상당한

상관관계 있음을 유추할 수 있는데, 고졸 이하 저학력 집단에서 동시에 나쁜 건강상태가 관찰될 가능성이 높다고 할 수 있다.

[그림 4]는 결혼이 가져오는 긍정적인 웰빙 효과만큼 이혼/사별의 부정적인 웰빙 효과도 잘 보여준다. 건강하다고 응답한 사람들만을 대상으로 하여 건강의 영향을 통제 한 경우에도 이혼과 사별을 경험한 집단에서 행복취약계층이 집중적으로 발생함을 알 수 있다. 이는 그룹 3 과 그룹 6 의 40, 50 대 중장년층에서 두드러지게 나타난다. 이는 이혼과 사별의 부정적인 웰빙 효과가 건강과는 독립적으로 특히 중장년층에 강하게 영향을 미치고 있음을 의미한다. 다만 이 경우에도 행복취약계층은 저소득층을 중심으로 발생하는데, 이는 소득과 무관하게 행복취약계층이 관찰되는 건강하지 않다고 응답한 집단과(그룹 7-그룹 12) 차별화되는 부분이다. 즉 건강이 뒷받침될 경우 이혼이나 사별이 가져오는 부정적인 웰빙 효과는 소득을 통해 어느 정도 상쇄될 수 있는 것으로 보인다.

## V. 결론과 함의

지금까지 논의는 데이터 시각화 기법을 통하여 우리나라의 행복취약계층은 어떠한 특징을 가지고 어떻게 분포하는지 살펴보았다. 이러한 시도는 두 가지 중요한 의미를 가진다. 우선 내용적으로 누가 행복취약계층에 속하고 이들이 어떤 특징을 가지고 있는지 이해하는 것은 우리나라의 평균적인 행복 수준을 향상시킬 행복정책을 디자인하는 출발점이 된다. 행복 수준을 무한하게 향상시킬 수 없고 행복 수준을 향상시키기 위한 자원의 투입이 가지는 한계효과는 감소한다는 점을 고려할 때, 행복취약계층은 행복정책의 주요 타겟이 될 수 있기 때문이다.

방법론적으로 본 연구에서 활용한 데이터 시각화 기법은 지금까지의 일반적인 통계적 방법론과는 상당한 차별점을 가진다. 최소한의 변수로 최대한의 분산을 설명한다는 간결성의 원칙에 기반한 일반적인 통계모형은 복잡하게 얽힌 사회현상을 설명하는데 한계가 있다. 가장 폭넓게 쓰이고 있는 회귀분석의 예를 들면 변수 간의 복잡한 관계를 설명하기 위해 사용하는 상호작용은 기껏해야 2 차항이며, 변수 간에는 기본적으로 선형관계가 가정된다. 그러나 사회현상은 이러한 회귀분석이 그럴 수 있는 세상 보다 훨씬 복잡하다. 데이터 시각화는 데이터가 가지고 있는 복잡한 정보를 있는 그대로 시각적으로 변환하여 제곱함으로 해서, 일반적인 통계모형에 비하여 정보 손실이 적고 패턴 및 추이를 파악하는데 용이하다. 복잡한 현상을 단순화하여 그 뼈대를 이해하는

데는 기존의 통계분석이 유리할 수 있지만, 복잡한 현상에 현미경을 들이대고 그 패턴과 추이를 보다 면밀하게 이해하는데 데이터 시각화 기법이 더 유용할 수 있다. 즉 두 기법은 어느 하나가 더 우월하기 보다 상호보완적인 성격이 강하며, 본 연구는 지금까지 많이 활용되지 않은 데이터 시각화 기법이 방법론적으로 어떤 유용성이 있는지 보여준다는데 의의가 있다.

본 연구의 분석결과는 연령대별로 어떤 변수가 행복취약계층과 관련되어 있고 따라서 행복취약계층은 어떤 특징을 가지는지 잘 보여준다. 우선 20, 30 대 청년층 중 행복취약계층은 실업/미취업 집단에 집중되어 있다. 특히 30 대의 경우 연령, 성별, 소득 등에 상관없이 취업 여부가 행복취약집단 분포에 결정적인 영향을 보이는 것으로 나타났다. 이는 청년층의 행복 결정에 일자리가 얼마나 중요한 영향을 끼치는지 잘 보여주는 결과이다. 이에 비해 40, 50 대 중장년층에서는 일자리와 함께 소득의 중요성도 부각되었다. 일자리가 있고 어느 정도의 소득활동이 이루어지게 되면 중장년층에서 행복취약계층으로 분류되는 경우는 매우 드문 것으로 보이기 때문이다. 다만 이러한 패턴은 주로 기혼집단에 국한되었고, 이혼/사별 집단의 경우 소득 최하위 집단 뿐 아니라 중간 소득 집단에서도 행복취약계층이 관찰되었다. 이는 중장년의 경우, 이혼/사별 등 결혼상태의 상실이 일종의 중요한 위생요인(행복수준을 낮추는 요인)으로 작용함을 의미한다. 60 대 이상 노년층에서 관찰되는 패턴은 행복에서 결혼이 가지는 중요성을 더욱 뚜렷하게 보여주었다. 노년층에서는 성별, 학력, 취업 여부 등 경제사회적 요인에 상관없이 혼인상태, 특히 이혼과 사별이 행복취약계층의 결정에 중요한 요인으로 작용하였다.

한편 기존 변수에 건강을 추가한 데이터 시각화 분석에서는 약간 다른 패턴이 관찰되었다. 건강은 기존의 다른 경제 사회적 요인들의 효과를 상당 부분 상쇄하고도 남음이 있었다. 행복취약계층은 건강하지 않은 집단에 집중되어 있는 곳으로 나타났고, 건강한 집단에선 일자리나 소득 등에 의한 행복취약계층의 의미 있는 차이가 관찰되지 않았다. 특히 건강이 뒷받침될 경우 중장년기 이혼이나 사별이 가져오는 부정적인 웰빙 효과도 소득 수준에 따라 어느 정도 극복 가능한 것으로 나타났다. 즉 건강은 다른 요인들에 우선하는 가장 기본적인 행복 결정요인으로 해석된다.

지금까지 연령, 성별, 소득, 취업여부, 결혼상태, 건강 등 다양한 경제 사회적 변수들을 중심으로 행복취약계층에 속하는 집단의 특징을 분석하였다. 정책의 측면에서 본 연구가 가지고 있는 시사점은 연령이나 혼인상태 등으로 세분화되는 집단에 따라 맞춤형 정책이 필요하다는 점이다. 20, 30 대 청년층 행복취약집단의 니즈와 40, 50 대 중장년층 행복취약집단의 니즈가 다르다. 40, 50 대 중장년층 행복취약집단 안에서도 여

성의 니즈와 남성의 니즈가 다르다. 따라서 국민 전체의 행복 수준을 높여야 한다는 규범적 담론만으로는 실제 우리 사회의 행복 수준을 향상시키는데는 한계가 있다. 다양한 변수에 따라 행복취약집단을 세분화하고, 이에 맞춘 세분화된 정책수단을 디자인하는 접근이야말로 행복중진국에서 행복선진국으로 발전하기 위한 필요조건이라고 할 수 있다.

## ▣ 참고문헌

- 구교준·임재영·최슬기. 2015. “행복에 대한 이론적 고찰.” 《정부학연구》, 21(2): 95-130.
- \_\_\_\_\_. 2017. 《무엇이 우리를 행복하게 하는가?》 서울: 집문당.
- 유정민·최영준. 2020. “복지국가는 어떻게 행복에 영향을 미치는가?: 자유안정성을 중심으로.” 《GRI 연구논총》, 22(1): 87-116.
- 장승진. 2011. “행복의 정치경제학: 소득불평등에 대한 인식이 한국인들의 삶의 만족도에 끼치는 영향.” 《한국정당학회보》, 10(2): 43-66.
- Becker, Gary S. 1975. *Human Capital: A Theoretical and Empirical Research*. New York: National Bureau of Economic Research.
- Bjornskov, Christian, Dreher, Axel, Fischer, Justina A. V., & Schnellenbach, Jan. 2009. “On the relation between income inequality and happiness: Do fairness perceptions matter?” *Center for European Governance and Economic Development Research, Discussion Paper* 91.
- Blanchflower, David G., & Oswald, Andrew J. 2004. “Well-being over time in Britain and the USA.” *Journal of Public Economics*, 88: 1359-1386.
- Clark, Andrew E., & Oswald, Andrew J. 1994. “Unhappiness and unemployment.” *Economic Journal*, 104: 648-659.
- Diener, Ed. 1984. “Subjective well-being.” *Psychological Bulletin*, 95: 542-575.
- \_\_\_\_\_. 1994. “Assessing subjective well-being: Progress and opportunities.” *Social Indicators Research*, 31: 103-157.
- Diener, Ed, Sandvik, Ed, Seidlitz, Larry, & Diener, Marissa. 1993. “The relationship between income and subjective well-being: Relative or absolute?” *Social Indicators Research*, 28: 195-223.
- Easterlin, Richard A. 1974. “Does economic growth improve the human lot?: Some empirical evidence.” In Paul A. David and Melvin W. Reder (eds). *Nations and Households in Economic Growth: Essays in Honour of Moses Abramovitz* (pp. 89-125), New York: Academic Press.
- \_\_\_\_\_. 2001. “Income and happiness: Towards a unified theory.” *The Economic Journal*, 111: 465-484.

- \_\_\_\_\_. 2003. "Explaining happiness." *Proceedings of the National Academy of Science*, 100: 11176-11183.
- Ferrer-i-Carbonell, Ada, & Ramos, Xavi. 2014. "Inequality and happiness." *Journal of Economic Surveys*, 28: 1016-1027.
- Frey, Bruno S., & Stutzer, Alois. 2000. "Happiness, economy and institutions." *Economic Journal*, 110: 918-938.
- Graham, Carol, & Chattopadhyay, Soumya. 2013. "Gender and well-being around the world." *International Journal of Happiness Development*, 1: 212-232.
- Hartog, Joop, & Hessel, Oosterbeek. 1998. "Health, wealth and happiness: Why pursue a higher education?" *Economics of Education Review*, 17: 245-256.
- Helliwell, John F., & Putnam, Robert D. 2004. "The social context of well-being." *Philosophical Transactions of the Royal Society*, 359: 1435-1446.
- Helliwell, John F., Layard, Richard, & Sachs, Jeffrey D. 2019. *World Happiness Report 2019*. New York: Sustainable Development Solutions Network.
- Helliwell, John F., Layard, Richard, Sachs, Jeffrey D., De Neve, Jan-Emmanuel, Aknin, Lara B., & Wang, Shun. 2022. *World Happiness Report*. United Nation.
- Inglehart, Ronald, Foa, Roberto, Peterson, Christopher, & Welzel, Christian. 2008. "Development, freedom, and rising happiness: A global perspective (1981-2007)." *Perspectives on Psychological Science*, 3: 264-285.
- Jeffreys, Harold. 1998. *The theory of probability*. Oxford: Clarendon Press.
- Kahneman, Daniel, & Krueger, Alan B. 2006. "Developments in the measurement of subjective well-being." *Journal of Economic Perspectives*, 20: 3-24.
- King, Gary, Keohane, Robert O., & Verba, Sidney. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press.
- Landes, Xavier. 2015. "Building happiness indicators: Some philosophical and

- political issues.” *The Ethics Forum*, 10: 4-37.
- Lane, Robert E. 2000. *The Loss of Happiness in the Market Democracies*. New Haven: Yale University Press.
- Layard, Richard. 2005. *Happiness: Lessons from a New Science*. London: Penguin Books.
- Mustonen, Pekka. 2009. “Health and Happiness: What’s Happening?” Finnish Medical Society Duodecim, Helsinki, Finland.
- Nussbaum, Martha. 2003. “Capabilities as fundamental entitlements: Sen and social justice.” *Feminist Economics*, 9: 33-59.
- Oishi, Shigehiro, Graham, Jesse, Kesebir, Selin, & Galinha, Iolanda Costa. 2013. “Concepts of happiness across time and culture.” *Personality and Social Psychology Bulletin*, 39: 559-577.
- Rothstein, Bo. 2010. “Corruption, happiness, social trust and the welfare state: A causal mechanisms approach.” *QoG Working Paper Series 2010:9*.
- Sen, Amartya. 1980. “Equality of what.” In Sterling M. McMurrin (ed). *The Tanner Lectures of Human Values* (pp. 257-280), Salt Lake City: Univ of Utah Press.
- Tokuda, Yasuharu, Fujii, Seiji, & Inoguchi, Takashi. 2010. “Individual and country-level effects of social trust on happiness: The Asia Barometer Survey.” *Journal of Applied Social Psychology*, 40: 2574-2593.
- Verme, Paolo. 2009. “Happiness, freedom and control.” *Journal of Economic Behavior and Organization*, 71: 146-161.
- Zagorski, Krzysztof, Evans, Mariah D. R., Kelly, Jonathan, & Piotrowska, Katarzyna. 2014. “Does national income inequality affect individuals’ quality of life in Europe? Inequality, happiness, finances, and health.” *Social Indicator Research*, 117: 1089-1110.
- Zweig, Jacqueline. S. 2015. “Are women happier than men? Evidence from the gallup world poll.” *Journal of Happiness Studies*, 16: 515-541.

## Finding happiness-vulnerable class using data visualization

Jun Koo & Dongwook Kim

The objective of this study was to identify a happiness-vulnerable class using data visualization techniques. Information loss is considered a major weakness of traditional linear models. It can be effectively overcome by data visualization. We combined two large social surveys from the Statistics Korea and the Korean Institute of Public Administration. These combined data had 3,360 sub-groups based on gender, age, education, income, health, employment status, and marital status. Data visualization technique effectively showed who were vulnerable groups with low happiness levels. Findings of this study imply that a more tailored policy approach according to different needs is needed.

※ Keywords: happiness, happiness-vulnerable class, data visualization

