

Stata를 이용한 OECD DAC 통계자료의 효율적 활용 방법 연구

이인호 (KOICA아카데미 ODA연구원 통계전문관)

제 I 장

제 II 장

섹
터
포
커
스

제 III 장

목차

I. 서론

II. 개발원조 데이터 개요

1. 한국의 개발원조 데이터 공개 현황
2. OECD 개발원조 데이터 개요
3. 원조 유형
4. 사업 분야
5. 전달 채널
6. 기타 분류 기준

III. OECD DAC 데이터를 위한 Stata 명령어

1. DAC 원시자료의 구조 및 주요 변수 설명
2. 주요 변수를 활용한 통계자료 작성

IV. Stata를 이용한 OECD DAC 데이터 활용 방법

1. DAC 원시자료의 패널데이터 구축
2. 시계열 통계자료 작성 및 패널 형태 변환

V. 결론 및 시사점

참고 문헌

요약

빅데이터가 개발협력 분야에서도 사용되기 시작하였다. 일부 공여국의 경우 개발협력의 특정 분야에서 빅데이터를 활용하고 있으며, 또 다른 공여국의 경우 앞으로 빅데이터를 사용하는 것에 대해 검토하고 있다. 따라서 향후 개발 문제를 해결하는 데 있어서 빅데이터의 수요는 늘어날 것으로 보이며, 이와 함께 데이터의 공급 및 활용 사례도 함께 증가할 전망이다. 그러므로 데이터를 통해 개발 문제를 다루기 원하는 데이터 사용자들은 이를 효과적으로 활용할 수 있는 역량을 갖추어야 할 것이다.

이러한 빅데이터의 확산과 활용 흐름에 맞추어 본 연구는 한국의 개발원조 데이터 공개 현황 및 한계, 경제협력개발기구(Organization for Economic Cooperation and Development, OECD)의 개발원조 데이터의 이해와 활용 방법을 살펴보았다. 또한 OECD의 대용량 데이터를 효과적으로 다루기 위해 통계 프로그램인 Stata의 명령어를 소개하며, 1973년부터 현재까지 누적된 300만 개 이상의 데이터를 효율적으로 활용할 수 있는 명령어(command)를 제시하였다.

따라서 본 연구는 개발협력 데이터를 사용하고자 하는 사용자의 공적개발원조(Official Development Assistance, ODA) 통계 지식의 제고뿐 아니라 연구 및 실무 분야에서 통계 작업의 효율성을 높이는 데 기여할 수 있을 것이다.

I. 서론

최근 들어 빅데이터(Big Data)가 개발협력 분야의 화두로 떠오르고 있다. 경제협력개발기구(Organization for Economic Cooperation and Development, 이하 OECD)의 데이터 관련 최근 보고서에 따르면 호주, 벨기에, 프랑스 등 9개의 회원국은 빅데이터가 개발협력 분야에 어떻게 기여할 수 있는지 이미 조사하기 시작하였고, 교육 및 보건 분야의 공공 정책을 개선하기 위해 빅데이터를 활용하고 있다. 또한 독일, 네덜란드, 영국 등 6개 회원국의 경우에는 아직 빅데이터를 도입하지 않았지만 개발 문제에 대한 이해, 모니터링 활동을 개선하기 위하여 빅데이터나 오픈 데이터를 사용했을 때의 장점에 대해 검토 중이다(OECD, 2017). 국제연합(United Nations, 이하 UN) 역시 2014년 제45차 UN 통계위원회(Statistical Commission)의 결정에서 빅데이터가 무시할 수 없는 정보의 원천(a source of information)임을 인정하였고, 지속가능목표를 달성하기 위한 빅데이터 사용을 위해 ‘빅데이터 글로벌워킹그룹(Global Working Group)’을 신설하였다.¹⁾

개발협력의 데이터와 관련해 국내 일부 선행연구에서도 개발 문제를 해결하기 위한 자료 확보와 이용자의 편의성 제고, 적극적인 데이터 활용을 강조하고 있다. 원시자료(raw data)는 프로젝트 평가 과정의 산출물로서 해당 프로그램의 성과 측정뿐 아니라 향후 다른 사업이나 정책 계획에도 유용하게 사용할 수 있다. 이를 위해 원조 기관은 원조 데이터를 체계적으로 관리하여 공개해야 하고, 데이터 사용자(data user)가 공개된 데이터를 적극적으로 활용할 수 있도록 관련 시스템도 구축해야 한다(김찬유, 2015). 특히 지속가능개발목표(Sustainable Development Goals, 이하 SDGs)의 이행과 모니터링을 위해 SDGs 지표 데이터의 경우, 지표의 이용 가능성 제고와 함께 ‘모든 지표, 모든 국가의 다년도 데이터’가 충분히 확보되어야 할 것이다(김수진, 2016).

이는 한국 원조에 대한 개발원조위원회(Development Assistance Committee, 이하 DAC)의 2017년 동료검토 결과와도 일치한다. ‘권고사항 2’ 중 ‘사업 예산의 규모 및 성과에 대한 정보 접근성 개선’을 위해 한국국제협력단(Korea International Cooperation Agency, 이하 KOICA)과 한국수출입은행 대외경제협력기금(Economic Development Cooperation Fund, 이하 EDCF)은 모두 성과관리 강화와 국제 수준에 맞는 정보 공개 기준을 수립하여 관련 자료를 공개하고 있다. 이러한 노력들은 2015년 한국 정부가 가입한 국제원조투명성기구(International Aid Transparency Initiative)의 취지에 맞게 앞으로도 지속적인 정보공개 확대로 이어져 원조 투명성 제고에 기여해야 할 것이다(조정화, 2018; 황재상, 2018).

1) <https://unstats.un.org/bigdata/> (접속일: 2018.05.18.)

위와 같이 개발협력 분야의 데이터 관련 이슈를 살펴보았을 때, 개발 문제를 해결하기 위한 빅데이터의 구축과 활용에 대한 수요와 공급이 늘어날 것으로 보인다. 아울러 현재 마주하고 있는 개발 문제를 해결하기 위해서는 질적연구와 함께 계량연구도 더욱 활발히 수행되어야 할 것이다. 이를 위해 데이터 사용자는 빅데이터의 구조와 체계(structure and system), 특히 개발원조 데이터의 항목(item or variable)과 의미(meaning)에 대한 정확한 이해와 활용 능력도 갖추어야 한다.

하지만, 빅데이터를 다루기에 앞서 데이터 사용자들은 공개된 많은 자료를 활용 가능한 자료(available data)로 가공하는 데 상당한 시간적 노력을 기울여야만 한다. 인터넷의 발달로 접근 가능한 자료(accessible data)는 많아졌지만, 데이터를 가공하고 변형하는 것(data munging)은 여전히 사용자의 몫으로 남아 있다. 결국 한정된 시간 안에 연구를 수행해야 하는 상황에서 사용자가 ‘자료를 자료답게’ 만드는 시간이 늘어난다면 연구의 질적 하락을 불러올 수 있으며, 연구의 효율성을 저해하는 요소가 될 수도 있다. 결국 계속 늘어나는 데이터를 다루는 데 있어서 데이터 사용자에게 ‘빠르고, 정확하며, 효율적인(prompt, accurate, and efficient) 자료 처리 도구(tool)’는 필수적인 요소가 될 것이다.

따라서 본 연구에서는 OECD DAC의 원시자료를 효과적으로 활용할 수 있는 방안에 대해 다룰 것이다.²⁾ 제2장에서는 한국의 개발협력 데이터 공개 현황과 한계점을 살펴보고, OECD 데이터 접근 및 이용 방법, 주요 항목의 의미를 설명할 것이다. 제3장에서는 1973년부터 현재까지 구축되어 있는 원시자료를 OECD 통계 홈페이지에서 다운로드하는 방법과 자료 처리를 위한 Stata 명령어³⁾를 소개할 것이다. 그리고 제4장에서는 제3장에서 다운로드한 연도별 횡단면 데이터(cross-sectional data)를 시계열 데이터(time-series data)로 병합하고, 패널 데이터(panel data)의 가로형(wide form) 및 세로형(long form) 변환 방법을 소개할 것이다. 마지막으로, 5장에서는 본 연구의 결론과 시사점을 제시할 것이다.

2) 세계은행(World Bank) 자료에 대한 효율적 활용 방안은 이인호(2018, 출간 예정) 참조.

3) Stata는 계량분석을 위해 사회과학 분야에서 널리 사용하는 통계 패키지 중의 하나로, 자료 분석뿐 아니라 명령문(command)을 통해 자료 처리도 빠르고 효율적으로 구현 가능함. 본 연구에서는 Stata 15.1 버전을 사용하였음.

II. 개발원조 데이터 개요

1. 한국의 개발원조 데이터 공개 현황

DAC 회원국은 매년 공적개발원조(Official Development Assistance, 이하 ODA)를 포함한 기타공적개발(Other Official Flows, 이하 OOF) 실적 등을 DAC 사무국으로 보고하고 있다. 통상 3월에는 잠정통계로서 전년도 실적 총액 및 요약테이블을 제출하고, 7월에는 사업별 세부 실적을 포함한 확정통계를 제출한다. DAC 사무국은 세부 실적을 취합한 이후 데이터 검증 거쳐 연말이나 다음 해 초에 OECD 통계 홈페이지에 원시자료를 포함하여 세부 실적을 공개한다.

한국 역시 2010년 DAC 가입 이후 위와 같은 절차에 맞추어 매년 통계 보고를 하고 있다. ODA를 집행하는 모든 부처 및 기관은 DAC에 해외 원조 실적을 보고하고 있으며, 한국수출입은행이 ODA 통계기관으로 지정되어 한국의 원조 실적을 취합하고 있다(국무조정실, 2010). 따라서 한국의 모든 ODA 통계정보는 한국수출입은행에서 운영하는 ‘ODA 통계조회시스템’⁴⁾을 통해 제공하고 있으며, 한국의 실적 중 KOICA의 통계정보는 ‘KOICA 통계조회서비스’⁵⁾를 통해 KOICA에서 제공하고 있다. 두 사이트 모두 다양한 항목별 분류를 통해 원조 유형별, 소득 수준별, 지역별, 국가별, 지원 형태별 등 사용자 편의에 맞게 쉽게 조회할 수 있는 메뉴로 구성되어 있다.

그러나 두 사이트 모두 통계 ‘조회’ 기능에 그쳐 있고 원시자료(raw data)의 제공 기능은 다소 취약하여 계량연구를 위한 분석자료로 사용하기에는 부족한 점이 있다. 더욱이 KOICA의 경우 통계조회서비스의 일부 메뉴(일반 검색, 상세 검색)에서 원시자료를 제공하지만, 자료의 양이 방대하여 자료를 다운로드하기 전에 우선적으로 조회를 하는 시간이 오래 걸린다. 또한 ‘공공데이터 포털’⁶⁾에서 KOICA ODA 통계정보의 원시자료를 일부 제공하지만, 모든 항목을 제공하지는 않고 오픈 API(application programming interface)로 자료를 얻어야 하는 등 일반인이 사용하기에는 다소 어려운 점 등이 있다.

4) ODA 통계조회시스템: https://stats.koreaexim.go.kr/index_outer.html (접속일: 2018.05.18.)

5) KOICA 통계조회서비스: <http://stat.koica.go.kr/> (접속일: 2018.05.18.)

6) 공공데이터포털: <https://www.data.go.kr/> (접속일: 2018.05.18.)

2. OECD 개발원조 데이터 개요

한국 ODA 자료의 원시자료를 앞 절에서 언급한 국내 통계 시스템을 통해 얻었다고 하더라도 실무나 계량연구 등의 분석에서 세계적 추세를 반영하기 위해서는 OECD나 세계은행(World Bank) 통계 사이트를 통해 자료를 얻어야만 한다. 이에 OECD DAC 통계 홈페이지에서는 1960년부터의 ODA 실적을 조회할 수 있으며, 원시자료의 경우에는 구분 기호(|, vertical bar delimiter)로 구분된 텍스트 파일(txt)을 1973년 실적부터 제공하고 있다. 이와 관련하여 OECD 홈페이지에서는 대량(bulk) 자료를 이용하려면 2007 버전 이상의 엑셀(Excel) 프로그램이나 다른 소프트웨어(Access, Stata 등)를 사용할 것을 안내하고 있다.⁷⁾

OECD에는 여러 분과위원회가 있는 만큼 다양한 분야의 통계를 제공하고 있다. 먼저 ‘OECD Data⁸⁾’에서는 농업, 개발, 경제 등 12개 주제별로 조회할 수 있다. 주로 차트, 그래프, 표 형태로 제공하며 사용자가 간단한 조건들을 변경해 가며 데이터를 확인할 수 있다. 예를 들어, 순(net) ODA 실적을 간단하게 확인하려면 아래와 같이 몇 번의 클릭만으로도 가능하다.

〈상자 1〉 OECD Data 접속 및 자료 확인 방법

OECD Data 접속(<https://data.oecd.org/>)
 ⇒ Topic에서 ‘Development’ 선택
 ⇒ Topic aspects에서 ‘Official development assistance(ODA)’ 선택
 ⇒ Development resource flows에서 ‘Net ODA’ 선택

출처: 저자 작성

위와 같이 조회하고자 하는 자료를 선택한 후에는 사용자의 필요에 따라 국가나 시간 범위를 재설정할 수 있다. 또는 그래프에서 원하는 국가만 강조(highlight)하여 쉽게 비교할 수도 있다. 이처럼 간단한 실적을 확인하는 데는 OECD Data 홈페이지가 유용하지만, 세부적이고 다소 복잡한 실적 조회는 다른 사이트를 이용해야 한다.

7) OECD: <http://www.oecd.org/dac/stats/idsonline.htm> (접속일: 2018.05.18.)

“Bulk downloads: Users can download entire datasets. Open your required dataset in OECD.Stat then from the Export menu, click on Related files. All .txt files can be opened in the Excel 2007 or later (use a vertical bar | as delimiter and ‘none’ as text qualifier) or any other software (for example Access, STATA, etc.) that allow more than 65,000 rows.”

8) OECD Data: <https://data.oecd.org/> (접속일: 2018.05.18.)

‘OECD.Stat’⁹⁾에서는 복잡한 데이터들을 표로 보여 줄 뿐 아니라 조회하는 표를 엑셀(Excel)과 같은 형태로 다운로드할 수 있으며, 나아가 관련 원시자료까지 제공한다. 앞에서 언급한 예제 데이터(net ODA)를 조회하기 위해서는 다음과 같은 절차를 따르면 된다.

〈상자 2〉 OECD.Stat 접속 및 자료 확인 방법

OECD.Stat 접속(<http://stats.oecd.org/>)
 ⇒ Data by theme에서 ‘Development’ 선택
 ⇒ 하위 메뉴에서 ‘Flows by Provider’ 선택
 ⇒ 하위 메뉴에서 ‘Total flows by donor(ODA+OOF+Private)[DAC1]’ 선택

출처: 저자 작성

조회 조건의 기본 값(default)으로 ‘Aid type’은 ODA, ‘Fund flows’는 Net Disbursements로, ‘Amount type’은 Current Price로 설정되어 있어 예제와 같은 데이터가 조회됨을 알 수 있다. ‘Donor’를 변경하기 위해서는 조회되는 표 안의 ‘Donor’ 글자를 클릭하여 체크 박스를 선택 혹은 해제할 수 있으며, 연도 설정 역시 ‘Year’ 글자를 클릭하여 같은 방식으로 변경할 수 있다. 이처럼 OECD.Stat에서는 조회할 수 있는 조건이 다양하지만, ODA 통계에서만 주로 사용하는 기본적인 항목들에 대한 설명은 다음과 같다.¹⁰⁾

3. 원조 유형

원조 유형(aid type)은 원조 전달(aid delivery)에 사용되는 원조 양식(aid modality)을 구분하는 것으로, 공여국의 자금이 어떠한 형태로 개발도상국에 이전되는지를 의미한다. 원조 양식은 총 8개(A~H)로 구분되어 있고, 하위 유형(sub type)까지 나타내면 다음 <표 1>과 같다.

9) 지금까지는 OECD 통계 홈페이지에 대한 구분을 하지 않았지만, 이후 사용하는 ‘OECD 통계 홈페이지’는 모두 OECD.Stat를 지칭함. (OECD.Stat: <http://stats.oecd.org/>) (접속일: 2018.05.18.)

10) 이후 3절부터 5절의 내용은 OECD DAC 통합통계보고지침(OECD, 2016a; OECD 2016b)과 이를 번역한 내용을 참고하여 작성함(한국국제협력단, 2016).

〈표 1〉 DAC 원조 유형 목록

A	A01	일반 예산지원
	A02	분야별 예산지원
B	B01	NGO, 시민사회, PPP, 연구기관에 대한 비지정 기여
	B02	다자기구에 제공하는 비지정 기여
	B03	국제기구(다자기구, INGO)의 특정 목적 프로그램 및 기금 지원
	B04	합동기금 지원
C	C01	프로젝트 원조
D	D01	공여국 전문가 / 봉사단 파견
	D02	기타 기술원조
E	E01	개발도상국 유학생 및 연수생 지원
	E02	연수생 교육기관 앞 간접지원비용
F	F01	채무 구제
G	G01	그 외의 항목에 포함되지 않은 행정성 경비
H	H01	개발인식증진
	H02	공여국 내 난민지원

출처: OECD (2016b), pp.75-80 내용을 저자가 재구성

현재 예시 화면에서 조회하는 각 원조 유형을 코드 값으로 표현하지는 않지만, 원조 유형은 위와 같이 고유 코드가 존재한다.¹¹⁾ DAC 통계보고지침(Statistical Reporting Directive, 이하 통계지침)(OECD, 2016a)에 따르면 각 원조 유형에 대한 의미는 다음과 같다.

A는 예산 지원(Budget support)으로서 본 범주에 속하는 자금 지원의 경우, 공여국은 수원국과 책임을 분담하여 해당 기금의 배타적 운영권을 포기해야 한다. 즉, 공여국의 지원 금액이 수원국의 예산으로 편입되는 경우가 A에 해당한다.

B는 비지정 기여와 합동 프로그램 및 기금(Core contributions and pooled programmes and funds)으로서 본 범주에 속하는 자금 지원의 경우, 공여국은 여타 이해 당사자들인 기타 공여국, 비정부기구(Non-Governmental Organization, 이하 NGO), 다자기구, 민관협력파트너십(Public-Private Partnership, 이하 PPP)과 책임을 분담하여 해당 기금의 배타적 운영권을 포기해야 한다. 또 하위 유형 중 B01(Core support to NGOs and civil society, PPPs and research institutes)과 B02(Core contributions to multilateral institutions) 그리고 B03(Contributions to specific-purpose programmes and funds managed by international organisations(multilateral, INGO)은 서로 그 의미가 다르다. B01과 B02는 각각 NGO와

11) 원조 유형에 대한 자세한 내용은 OECD(2016b)의 “Annex11. 원조 유형과 구축성 상태” 참조.

다자기구에 대한 비지정기여(core contribution)로, 각 기관이나 기구의 예산으로 편입되는 경우다. 이때 공여국은 ODA 적격 국제기구 목록(Annex2)¹²⁾에 포함된 곳에만 지원할 수 있다. B02의 경우 양자·다자 구분에서 ‘다자’로 구분하며, Annex 2에 있는 비지정기여 계수(coefficient for core contribution)¹³⁾에 따라 ODA 지원 금액이 결정된다. 그러나 B03은 B02와는 달리, 비지정기여가 아닌 국제기구가 관리하는 특정 목적의 프로그램이나 기금에 제공하는 공여국의 지원을 의미한다. 이를테면 ‘유니세프 여성 교육’, ‘아프가니스탄 재건 신탁기금’과 같은 프로그램 또는 신탁기금 등이 이에 해당한다.

C는 프로젝트 원조(Project-type interventions)로서 지정된 시한과 예산, 지역을 대상으로 특정한 목표와 성과를 달성하기 위해 파트너 국가와 협의한 일련의 투입, 원조 활동 및 산출물로 이루어진 사업이다. 여기서 유의할 것은 어떤 프로젝트가 NGO나 다자기구를 ‘통해(through)’ 수행된다면 B01이나 B02가 아닌 C01로 기록한다. 앞에서 언급한 바와 같이 B01, B02로 기록되는 지원은 NGO나 다자기구 ‘에(to)’ 제공하는 원조다. 이때 공여국의 재정 운영권은 없으며, 해당 지원은 NGO나 다자기구가 재량에 따라 사용할 수 있는 공공 기금 지원이다. 반면 공공 분야에서 지정 혹은 인지하거나 승인한 목적과 관련된 지원, 즉 지원의 사용처가 표시된(earmarked) 지원의 경우에 NGO나 다자기구는 공공 분야를 ‘대신하여’ 공공자금을 사용하는 것이므로 해당 사업은 C01로 기록한다.

D는 전문가와 기타 기술지원(Experts and other technical assistance)으로서 본 범주에는 인력, 훈련 및 연구의 형태로 노하우를 제공하는 외부 프로젝트가 포함된다. KOICA의 경우 봉사단과견사업은 D01(공여국 전문가 파견 비용)로, 개발도상국 연수생초청사업은 D02(기타 기술원조)로 분류한다.

E는 공여국 내 장학금과 유학생 지원비용(Scholarships and student costs in donor countries)이다. KOICA의 개발도상국 석사과정지원사업은 E01(공여국 내 장학금 및 훈련비용)에 해당한다.

F는 채무구제(Debt relief)로서 채무탕감, 전환, 스와프, 환매, 재조정, 재융자 등 채무와 관련한 모든 활동을 포함한다. 또 G는 그 외의 항목에 포함되지 않은 행정성 경비(Administrative costs not included elsewhere)이고, H는 기타 공여국 내 지출(Other in-donor expenditures)로서 국가 간 자금흐름을 초래하지 않는 다수의 자금지원 형태가 여기에 해당한다.

12) ODA 적격 국제기구 목록은 OECD(2016b)의 “Annex2. ODA 적격 국제기구 목록” 참조.

13) 해당 기관의 예산 중 순수하게 ODA 정의에 부합하는 지출 금액의 비율을 의미하며, DAC 사무국이 통계작성반의 협의를 거쳐 설정함.

4. 사업 분야

앞에서 예시로 설명한 조회 페이지(DAC1)는 가장 기본적인 정보들만 있으므로, 이외에 더욱 세부적인 데이터를 확인하려면 다른 메뉴를 조회해야만 한다. 우선, Development 하위 메뉴 중 ‘Flow based on individual projects(CRS)’를 선택하고, 하위 메뉴에서 ‘Creditor Reporting System(CRS)’을 선택한다. 이 페이지에서는 앞에서 조회했던 것보다 더 많은 조건을 볼 수 있다.

본 페이지에서는 원조 유형 외에도 해당 사업의 분야(sector) 정보도 제공하고 있다. 통계지침에서 분야, 즉 원조목적(purpose of aid)은 공여국의 자금을 통해 수원국의 경제 또는 사회구조의 어떤 분야를 발전시키고자 하는가에 대한 답변으로 생각해 볼 수 있다. 따라서 공여국의 지원 내역은 가능한 한 가장 구체적인 코드로 지정되어야 하며, 다음 <표 2>의 넓은 범주에 속하는 세부 코드를 선택해야 한다(OECD, 2016a).

<표 2> DAC 분야 목록

사회 인프라와 서비스	교육, 보건, 인구, 수자원, 정부와 시민사회 분야
경제 인프라와 서비스	교통, 통신, 에너지, 금융 및 재무, 비즈니스 서비스
생산	농업, 임업, 어업, 산업, 광업, 건설, 무역, 관광
다분야/범분야	일반 환경보호, 도시·농촌 개발을 포함한 기타 다분야
비배분분야	일반 예산지원, 채무 관련 활동, 인도적 지원, 공여국 내 자금 이전과 같이 분야별 배분이 어려운 자금지원

출처: OECD(2016b), pp.84–107 내용을 저자가 재구성

위 DAC 분야 목록에서 세부 분류 항목은 5자리 코드 체계를 나타내고 있으며, Creditor Reporting System의 앞 글자를 따서 일반적으로 CRS 목적코드(CRS purpose code)¹⁴⁾로 부른다. CRS 코드 5자리 중 앞 3자리는 분야를, 뒤 2자리는 기능을 의미한다. 이를테면 121은 ‘보건일반’ 분야를, 151은 ‘공공행정 및 시민사회’ 분야를 나타내며, 여기에 ‘정책 및 행정’ 기능을 나타내는 10이 결합할 수 있다. 즉, 12110은 ‘보건정책 및 행정관리’, 15110은 ‘공공정책 및 행정관리’를 의미한다. 전체 CRS 코드는 필요에 따라 신설하거나 통폐합하는 등 비정기적으로 개정하고 있다(OECD, 2016b).

14) CRS 목적코드에 대한 자세한 내용은 OECD(2016b)의 “Annex12. 분야별 목적과 목적코드 분류” 참조.

5. 전달 채널

원조 유형과 분야 외에 채널(channel) 역시 원조 관련 데이터를 구분할 수 있는 분류 중 하나다. 전달 채널(channel of delivery)이란 최초의 원조시행 파트너로서 원조 시행의 의무를 가지고 통상적으로 시행 기관과 계약 또는 기타 구속력 있는 협정으로 맺어져 있으며, 기금에 직접적인 책임을 진다. 이를 통해 특정 다자기구에 제공하는 비지정기여를 파악할 수도 있으며 또 다자기구, NGO, PPP 및 기타 채널을 통해 제공한 양자 간 원조의 총액 산출도 가능하다. 또한 공공·민간 원조시행파트너를 구분할 수도 있다(OECD, 2016a).

이를테면, KOICA가 어떤 개발도상국에 병원을 건립하는 프로젝트(C01)를 수행한다고 할 때 전달 채널은 공공분야기관(10000)이다. 반면, 동일한 사업(C01)을 다자기구가 수행한다고 할 때 전달 채널은 다자기구(40000)가 된다. 전달 채널 역시 조회 화면에서는 코드 값이 보이지 않는다. 하지만, 아래 <표 3>과 같이 고유한 코드를 가지고 있다.¹⁵⁾

〈표 3〉 DAC 원조전달채널 목록

10000	공공분야기관
20000	NGO 및 시민사회단체
30000	PPP 및 네트워크
40000	다자기구
50000	기타
60000	민간분야기관
90000	기타 ¹⁶⁾

출처: OECD (2016), pp.69-72 내용을 저자가 재구성

<표 3>은 원조전달채널의 상위 코드만 나타낸 것으로, 7개의 상위 코드 중 일부는 중간분류 코드와 특정 기관마다 부여된 고유 코드로 나뉘어 있다.

우선 10000의 경우에는 공공분야기관을 나타내며, 11000은 공여국 정부, 12000은 수원국 정부, 13000은 제3국 정부를 의미한다. 세부적으로는 중앙정부, 지방행정, 공기업, 공여국 내 기타 공공기관으로 분류된다. KOICA의 경우, 공여국 정부를 의미하는 11000 코드를 사용하고 있다.

15) 전달 채널에 대한 자세한 내용은 OECD(2016b)의 “Annex 9. 주요 원조전달채널” 참조.

16) 대학교, 단과대학, 기타 교육기관, 연구소 또는 싱크탱크 등의 기타를 나타내는 50000에도 해당하지 않는 기타 채널을 의미함.

20000은 NGO를 나타내며, 21000은 국제 NGO, 22000은 공여국 기반 NGO, 23000은 수원국 기반 NGO로 분류한다. 일부 국제 NGO는 고유 코드를 가지고 있으며, 21000의 하위 코드로서 국경없는의사회에는 21029, 국제적십자위원회는 21016이 부여되어 있다.

30000은 PPP 및 네트워크를 의미하며, 31000은 PPP, 32000은 네트워크를 의미한다. 20000과 마찬가지로, 일부 특정 PPP 및 네트워크에는 30000의 하위 코드로서 고유 코드가 부여되어 있다.

40000은 다자기구이며, 41000은 UN, 42000은 유럽연합(European Union, 이하 EU), 43000은 국제통화기금(International Money Fund, IMF) 등으로 분류되어 있다. 마찬가지로, 특정 기관은 세부 코드가 부여되어 있으며, 유엔개발계획(United Nations Development Programme)은 UN의 41000 코드 아래 41114, 유럽투자은행(European Investment Bank)은 EU의 42000 코드 아래 42004다.

50000은 기타로서 대학교, 단과대학, 기타 교육기관, 연구소 또는 싱크탱크 등을 포함한다. 또 60000에는 공여국, 수원국, 제3국의 민간 분야 기관이 포함되어 있다. 역시 특정 기관에는 고유 코드가 부여되어 있다.

6. 기타 분류 기준

앞에서 설명한 원조 유형, 분야, 전달 채널 외에도 CRS 조회 페이지에는 Donor, Recipient, Flow, Amount type, Flow type 등의 항목이 있다. Donor에서는 공여국을 선택할 수 있으며, DAC 회원국 이외에 다자기구와 비회원국의 공여 실적까지 확인할 수 있다. 또한 Recipient에서는 수원국을 선택할 수 있으며, 현재까지 적격 수원국 목록에 등재된 적이 있는 국가는 모두 조회 가능하다.¹⁷⁾ 그리고 Flow에서는 자금의 성격에 관한 것으로서 ODA, OOF 그리고 민간증여(Private Grant)로 나뉘어 있다. 아울러 Amount type은 물가변동 기준의 반영 여부에 대한 선택이며, 보고 당시의 가격인 current price와 물가 변동을 반영한 constant price로 조회할 수 있다. 또 Flow type에서는 약정액과 지출액을 선택할 수 있다.

지금까지 언급한 내용들은 DAC 데이터를 분석하기 전에 연구자가 기본적으로 숙지해야 하는 분류다. 원시자료에는 위에서 설명한 세부 코드들을 포함하고 있으며, 원시자료를 사용하기 전에 각 코드의 개념과 의미를 명확히 알아야 통계의 오류를 줄이고 또 신뢰도 높은 통계자료 가공 및 분석을 할 수 있다.

17) 수원국 목록에 대한 자세한 내용은 OECD(2016b)의 “Annex 8. 수원국 목록” 참조.

III. OECD DAC 데이터 활용을 위한 Stata 명령어

본 장에서는 2장에서 조회하였던 작업 외에 Stata에서 원시자료를 활용할 수 있는 명령어의 사용법에 대해 소개할 것이다. OECD 통계 홈페이지는 조회 화면 외에 원시자료를 다운로드하는 기능을 제공하고 있다. 다운로드 경로는 아래와 같다.

〈상자 3〉 OECD DAC 원시자료 다운로드 방법

OECD.Stat 접속(<http://stats.oecd.org/>)
 ⇒ Data by theme에서 'Development' 선택
 ⇒ 하위 메뉴에서 'Flows based on individual projects(CRS)' 선택
 ⇒ 하위 메뉴에서 'Creditor Reporting System(CRS)' 선택
 ⇒ 상단 메뉴에서 'Export' 선택 후 'Related files' 선택
 ⇒ 'Ready-made files' 팝업 박스에서 관련 파일 클릭하여 다운로드

출처: 저자 작성

현재 OECD 통계 홈페이지에서는 1973년부터 2016년까지의 원시자료를 게시하였고, 2018년 4월 9일 기준으로 모든 파일이 업데이트되었다. 본 장에서는 모든 파일(16개)이 Stata 작업 폴더(working directory)에 있다는 전제하에 설명할 것이다.¹⁸⁾

1. DAC 원시자료의 구조 및 주요 변수 설명

우선, 모든 파일을 다운로드한 후 압축을 풀어 텍스트 파일(*.txt)을 Stata 작업 폴더(예시: c:\myfolder\)로 이동시킨다. 해당 파일은 각각의 값을 구분 기호(|)로 구별하게끔 저장된 파일이다. 엑셀(Excel)에서 사용하고자 할 때는 해당 파일을 먼저 불러온 후, 텍스트 나누기 기능에서 구분 기호를 '|'로 선택해서 변환해야 한다. Stata에서는 import 대화창을 실행시키거나 명령어(command) 창에서 다음과 같은 명령어로 파일 불러오기(import)를 수행할 수 있다.

```
. import delimited "CRS 2016 data.txt", delimiter("|")
```

예제 코드에서는 2016년 자료만 Stata로 불러왔으며, 실행 결과를 보면 86개의 변수와 약 24만 개의 관측치가 잘 불러 온 것을 확인할 수 있다. 데이터를 보는 명령어(browse 또는 br)로 익숙한 엑셀(Excel) 화면과 같이 조회할 수도 있다.

18) 3장과 4장에서 사용한 몇 가지 Stata 명령어는 민인식·최필선(2015, 2016)과 Stata 도움말을 참고하였음.

DAC 원시자료는 공여국(기관)이 보고하는 한 줄, 한 줄이 모두 취합되어 1년의 자료가 완성된다. 이는 특정 시점의 원조 관련 변수가 각 공여국 및 수원국 등으로 잘 정리되어 있는 횡단면 데이터(cross-sectional data)다. 각 변수들을 확인하기 위해서는 command 창에 codebook을 입력하여 모든 변수의 자료 특성을 확인할 수 있다. 본 연구에서는 몇 가지 중요한 변수에 대해서만 확인하고자 다음과 같이 입력한다.

```
. codebook year donorname recipientname flowname aid_t usd_disbursement purposecode channelcode
```

〈그림 1〉 일부 변수에 대한 codebook 실행 결과

year		
type:	string (str4)	
unique values:	2	missing "": 0/224,373
tabulation:	Freq. Value	
	1 "→"	
	224,372 "2016"	
donorname		
type:	string (str53)	
unique values:	91	missing "": 1/224,373
examples:	"EU Institutions"	
	"International Bank for Reconstruction and Development"	
	"Luxembourg"	
	"UNICEF"	
warning:	variable has embedded blanks	

출처: 저자 작성

〈그림 1〉에 나타난 것과 같이 year는 연도를 의미하며, 현재 2016년 데이터만 불러왔으므로 값(Value)은 '2016'만 있어야 하지만, 1개 값에는 화살표('→')가 입력되어 있는 것을 확인할 수 있다. 이는 각 원시자료의 가장 마지막에 데이터가 붙어 있는 표시로, br을 통해 확인 가능하다. type에서는 2016이 숫자지만 화살표로 인해 데이터 형태(type, 이하 type)가 문자형(string, 이하 string)으로 저장된 것을 확인할 수 있다. 다음과 같이 연도값이 2016이 아닌 값은 삭제하도록 입력하여 원 자료와 상관없는 값을 제거할 수 있다.

```
. drop if year != "2016"
. destring year, replace
```

실행 결과 1개의 관측치가 삭제된 것을 확인할 수 있고, 다시 codebook year를 실행하면 모든 값이 '2016'인 것을 확인할 수 있다. 문자를 숫자로 변경해 주는 명령어(destring)도 실행하였으므로, type이 숫자형(numeric(int))으로 변경된 것도 확인 가능하다.

최초 수행한 codebook 명령어의 결과값을 중심으로 각 변수를 다시 설명하면, donorname과 recipientname은 각각 공여국과 수원국의 명칭으로서 모든 국가 명에는 donorcode, recipientcode 변수에 있는 특정 코드가 부여되어 있다. type은 모두 string이며, 앞에서 설명한 의미 없는 화살표값으로 인해 결측치(missing, 이하 missing)가 1건 있음을 알 수 있다.

flowname은 ODA의 무상(grants), 유상(loans) 그리고 OOF 등을 구분하는 변수로서 unique values를 통해 5종류의 고유 값이 있음을 확인할 수 있고, tabulation(또는 tab) 명령어로 각 변수값의 예시를 확인할 수 있다.

aid_t는 원조 유형(aid type)을 의미한다. 데이터 예시는 'C01'과 같은 값이 입력되어 있고, type은 string, missing은 768개가 있음을 알 수 있다. 아래와 같이 입력하면 missing이 어디에 있는지 대략적으로 확인할 수 있다.

```
. tab flowname if aid_t == ""
```

실행 결과 일부 ODA 무상 및 유상, OOF에서 원조 유형에서 missing이 있는 것을 알 수 있다(<그림 2> 참조).

〈그림 2〉 tabulate 실행 결과 1

FlowName	Freq.	Percent	Cum.
ODA Grants	15	1.96	1.96
ODA Loans	49	6.39	8.34
Other Official Flows (non Export Cred..	703	91.66	100.00
Total	767	100.00	

출처: 저자 작성

제I장

제II장

섹
터
포
커
스

제III장

usd_disbursement는 각 사업별 미화 지출액을 의미하며, 단위는 백만 달러다. type이 숫자 형이므로 최솟값과 최댓값이 range에 나와 있고, 평균(mean)과 표준편차(std. dev)도 제시하고 있음을 확인할 수 있다.

purposecode는 사업 분야의 CRS 목적코드를 의미한다. 원래 5자리지만, range를 살펴보면 100부터 값이 있음을 알 수 있다. 아래와 같이 입력하여 어떤 값들이 있는지, 또 어떤 데이터가 3자리만 부여되어 있는지 확인할 수 있다.

```
. tab purposecode if purposecode < 1000
. tab flowname if purposecode < 1000
```

〈그림 3〉 tabulate 실행 결과 2

FlowName	Freq.	Percent	Cum.
Other Official Flows (non Export Cred..	702	100.00	100.00
Total	702	100.00	

출처: 저자 작성

실행 결과를 보면 <그림 3>과 같이 702개의 값이 3자리 코드로 입력되어 있고, 모두 OOF에 분포해 있는 것을 알 수 있다.

위에서 실행한 codebook 명령어의 마지막인 channelcode는 전달 채널로서 type은 숫자, 10000부터 90000까지 275개의 고유값, 약 16,000개의 missing이 있음을 확인할 수 있다. 변수가 숫자이므로 평균과 표준편차, 분위수(10, 25, 50, 75, 90)가 제공되지만 사실상 의미는 없다.

위 변수들은 OECD 통계 홈페이지에서 기본적으로 제공하는 조회 서비스에 포함되어 있지만, 사실 소개하지 않은 변수들을 다양하게 가공하여 사용자가 원하는 정보들을 산출할 수 있다. 이를테면 공여국별 평균지출금액, 수원국의 원조 유형별 평균지출금액 등의 정보를 가공하는 것도 가능하다. 또한 원시자료에는 각 공여국의 공여기관(agencycode, agencyname)에 대한 정보도 있으므로 특정 기관의 사업 정보를 조회할 수도 있다.

2. 주요 변수를 활용한 통계자료 작성

앞에서 설명한 주요 변수들을 통해 원시자료에서는 OECD 통계 홈페이지에서 보여 주지 못하는 통계자료를 작성할 수 있다. 먼저, 금액과 관련된 부분이다. OECD 통계 홈페이지에서는 각 필드별 합계 금액만 보여 주지만, 원시자료를 가공하면 평균, 표준편차, 최소 및 최댓값도 확인이 가능하다.

```
. sum usd_disbursement if flowname == "ODA Grants" & bi_multi == 1
```

〈그림 4〉 summarize 실행 결과

Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	144,773	.8250923	24.99939	0	6585.08

출처: 저자 작성

〈그림 4〉를 살펴보면, 지출액에 대해 summarize(또는 sum) 명령어를 수행하면 간단한 산술통계치를 확인할 수 있다. 조건은 무상원조(ODA Grants)와 양자(bi_multi=1) 중에서만 확인하는 것으로 규정하였다. 대략적인 의미는 2016년의 1개 사업당 약 백만 달러, 최솟값은 0, 최댓값은 약 6,585백만 달러다. 다음 명령어를 통해 지출액이 가장 많은 사업은 어떤 것인지 확인할 수 있다.

```
. gsort -usd_disbursement
. list donorname aid_t projecttitle usd_disbursement in 1
```

위 명령어는 우선 금액 기준으로 해서 내림차순(-)으로 정렬(gsort)한 후, list 명령어로 공여국, 원조 유형, 사업 명, 금액의 첫 번째 줄(in 1)을 보여 주는 것이다. 확인 결과, 독일의 공여국 내 난민 지원 금액임을 알 수 있다.

국가별 산술 통계를 보여 주는 것도 가능하다. 한국과 일본의 프로젝트 사업(C01)의 평균 금액을 확인하고자 할 때는 다음과 같은 명령어를 사용한다. 우선, 현재 전체 공여국에 대한 데이터를 메모리에 저장(preserve)한 후, 공여국 명(donorname)이 한국(Korea)인 경우나 일본(Japan)인 경우만 남기고(keep) 모두 삭제한다. 이후 공여국별로 접두사(prefix) 명령어 bysort:를 통해 프로젝트(C01)인 경우의 지출액을 조회한다. 이후에는 메모리에 저장한 데이터를 불러오는 restore 명령어를 사용하였다.

```
. preserve
. keep if donorname == "Korea" | donorname == "Japan"
. bysort donorname: sum usd_disbursement if aid_t == "C01"
. restore
```

〈그림 5〉 bysort 실행 결과 1

-> donorname = Japan					
Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	10,662	.9319335	8.318833	9.19e-06	476.942
-> donorname = Korea					
Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	1,146	.9544534	2.566816	0	30.4801

출처: 저자 작성

〈그림 5〉의 자료를 해석하면 1개 사업당 평균 금액은 약 1백만 달러로 비슷하지만, 사업 개수는 일본이 더 많고, 최댓값과 최솟값의 범위가 넓어 분산 역시 일본이 한국보다 더 크다는 것을 알 수 있다.

또한 원시자료에서는 공여국 안의 공여 기관별 원조 정보를 확인하는 것 역시 가능하다. 공여국 관련 변수(donorname, donorcode) 외에 기관 관련 변수(agencyname, agencycode)를 활용할 수 있다. 공여국과 마찬가지로 각 국가별 공여 기관에도 고유 코드가 부여되어 있으며, 다음과 같은 코드를 통해 확인 가능하다. 예시에서는 영국 개발협력청(Development for International Development)의 프로젝트(C01)와 기타 기술협력(D02) 지원액을 확인하였다.

```
. preserve
. drop if agencycode != 1 | donorcode != 12
. keep if aid_t == "C01" | aid_t == "D02"
. bysort aid_t: sum usd_disbursement
. restore
```

〈그림 6〉 bysort 실행 결과 2

-> aid_t = C01					
Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	2,208	2.195198	8.744068	0	282.272
-> aid_t = D02					
Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	264	1.481154	3.929004	0	48.3274

출처: 저자 작성

실행 결과, <그림 6>과 같이 C01은 총 2,208개, D01은 264개이며, 각각의 평균 사업비는 약 220만 달러와 148만 달러로 차이가 있음을 알 수 있다. 각 원조 유형에 대한 분산과 최솟값 및 최댓값도 확인 가능하다.

제I장

제II장

섹
터
포
커
스

제III장

IV. Stata를 이용한 OECD DAC 데이터 활용 방법

지금까지 1개 연도의 자료를 이용하며 DAC 원시자료에 대한 간단한 활용 방법을 살펴보았다. 그러나 실제 자료를 활용할 때는 횡단면 데이터만 사용하는 것이 아니라 시간 개념이 들어간 시계열 데이터(time-series data) 및 패널 데이터(panel data)를 사용하기도 한다. 따라서 본 장에서는 연도별로 다운로드한 DAC 원시자료를 1개의 파일로 병합하고, 3장에서 수행한 예제 코드를 확장하여 자료를 분석하는 방법을 소개할 것이다. 3장과 마찬가지로 Stata의 작업 폴더는 여전히 myfolder(예시: c:\myfolder\)다.

1. DAC 원시자료의 패널 데이터 구축

다운로드한 연도별 원시자료를 1개의 파일로 만들기 위해서는 아래와 같은 명령어를 수행한다. 지역변수(local)를 활용하는 매크로의 개념이 들어가므로 do file editor에서 수행하는 것이 편리하다.

```
local txtfiles: dir . files "*.txt"
foreach files of local txtfiles {
    import delimited "files", delimiter("|") clear
    local time = c(current_time)
    local temp = subinstr("time",".", "_", ..)
    save `temp'
}

local dtfiles: dir . files "*.dta"
local temp1 : word 1 of `dtfiles'
use "temp1", clear
rm `temp1'

local dtfiles: dir . files "*.dta"
foreach files of local dtfiles {
    append using "files", force
}

save DacPanelData
```

먼저, 작업 폴더에서 원시자료의 파일(txt 파일)을 txtfiles이라는 local로 저장한다. 그리고 foreach 반복문을 통해 local값, 즉 파일 명을 3장에서 불러온 것과 같은 방식(import delimited)으로 순차적으로 수행하여 현재 시각(시:분:초)을 파일 명으로 임시 저장(dta 파일)한다. 다음에는

임시로 저장한 dta 파일을 다시 불러오는데, 첫 번째 파일만 Stata에서 불러온 후 삭제하고, 이후 반복문을 통해 나머지 파일을 모두 병합(append)한다. 이후 완성된 파일(DacPanelData.dta)을 저장하고 종료하면 1973년부터 2016년까지 DAC 원조데이터가 1개의 파일로 생성하였고, 용량이 약 7GB인 Stata 데이터 파일(dta)을 확인할 수 있다.

약 20분 정도 시간이 소요되지만,¹⁹⁾ 대용량의 연도별 데이터를 엑셀(Excel)에서 연구자가 일일이 ‘텍스트 나누기’부터 시작하여 파일 병합(copy and paste)까지 수행하는 것에 대한 번거로움과 피로를 고려하면 Stata를 이용하는 것이 효율적이고 효과적인 대안이다. 설령, 엑셀(Excel)에서 이 작업을 수행했다 하더라도 엑셀(Excel)의 메모리가 대용량 데이터를 관리할 수가 없어 프로그램이 제대로 작동하지 않을 것으로 보인다. 또한 엑셀(Excel) 2013 기준으로 관리할 수 있는 데이터의 수는 약 104만 개에 불과하다. 더욱이 DAC 원시자료에 있어 2016년도까지의 관측치가 약 350만 개를 넘고 또 최근 추세를 보면 매년 약 20만 개씩 추가될 것으로 보이므로, 사실상 엑셀(Excel)에서는 불가능한 작업이다.

tab year 명령어를 통해 데이터를 살펴보면 의미 없는 화살표 데이터가 있는 것을 알 수 있고, 1970년부터 2016년까지 빠짐없이 데이터가 있는 것을 확인할 수 있다. 자료를 분석하기 전에 3장에서 수행했던 것과 같이 화살표를 제외한 자료만 남겨 둔다. 명령어는 다음과 같이 공여국 코드(donorcode)의 missing을 제거하게끔 작성하여 수행할 수 있다.

```
. drop if donorcode == .
```

2. 시계열 통계자료 작성 및 패널 형태 변환

1개의 파일에 다년도의 데이터가 있으므로, 이제는 3장에서 수행한 명령어들을 기반으로 시계열 통계자료를 작성할 수 있다. 예시에서는 우선 다음과 같이 연도변수(year)를 숫자화(destring)한 후, 편의상 자료를 최근 3년으로 한정하고 일부 변수를 삭제하도록 하겠다.

```
. destring year, replace
. keep if year >= 2014
. drop crsid-initialreport regioncode-incomegroupname usd_received-usd_grantequiv
  geography-capitalpend
```

19) Stata/SE 15.1, i7 3.40GHz, 8GB 기준. dta 임시 파일(temp)을 저장하지 않고, 1개의 foreach 반복문에서 연도별 파일을 메모리에 저장하고 불러들이면서(preserve, restore) 코딩할 수도 있다. 그러나 연산 시간을 계산한 결과, 용량이 커짐에 따라 메모리를 많이 사용하여 약 30분 정도 소요되어 2개의 foreach 반복문으로 코딩함.

우선, 연도별 지출액을 살펴보기 위해 다음과 같이 bysort를 사용하면 <그림 7>과 같은 결과를 얻을 수 있다.

```
. bysort year: sum usd_disbursement
```

〈그림 7〉 연도별 자료의 bysort 실행 결과

-> year = 2014					
Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	211,105	1.064847	12.08951	-37.4429	1879.03
-> year = 2015					
Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	215,150	1.156678	22.65229	-24.263	5310.24
-> year = 2016					
Variable	Obs	Mean	Std. Dev.	Min	Max
usd_disbur~t	211,404	1.171172	23.34521	-109.697	6585.08

출처: 저자 작성

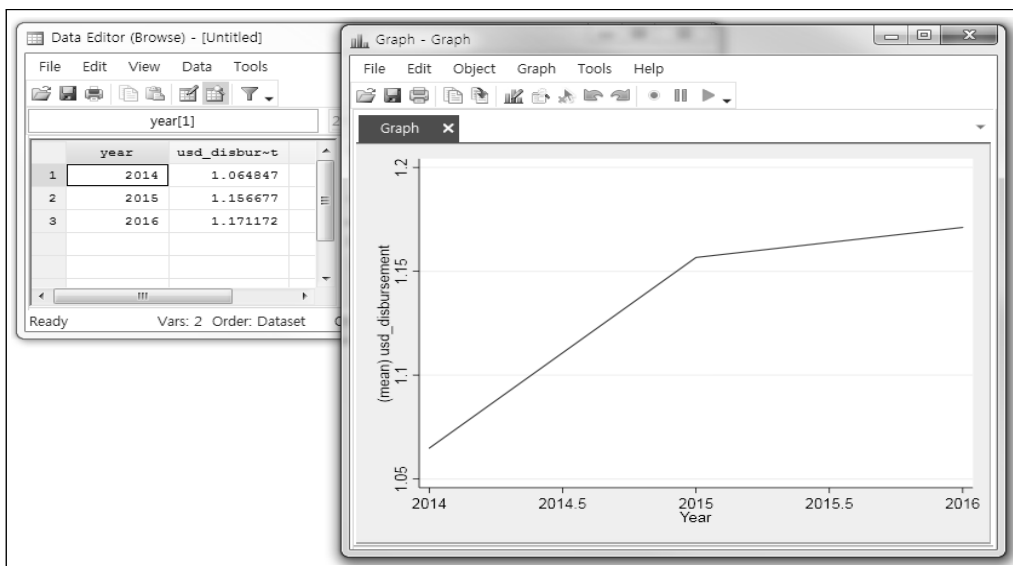
평균적으로 사업비의 최댓값은 연도가 지날수록 증가하고 있으며, 사업비의 분산 역시 증가하는 것으로 나타났다. 그러나 이런 시계열 자료를 사용하기 위해 Stata의 command 결과 창에서만 본다면, 자료로 활용하기가 매우 불편하다. 따라서 엑셀(Excel)의 pivot 기능과 유사한 collapse 명령어와 함께 reshape 명령어를 사용하여 가로형(wide form) 또는 세로형(long form) 패널 데이터로 구현할 것이다.

먼저, collapse 명령어는 데이터를 축약하는 명령어로서 필요한 통계치와 변수를 선택하여 보여 준다. 아래와 같이 입력하면 지출액의 평균을 연도별로 보여 주는 간단한 데이터 테이블이 생성된다. 예시 자료를 계속 활용할 것이므로, preserve와 비슷한 메모리 임시 저장 명령어 snapshot으로 데이터를 저장하기로 한다.

```
. snapshot save
. collapse usd_disbursement, by(year)
. line usd_disbursement year
```

collapse에 아무런 조건도 주지 않으면 기본적으로 평균(mean)을 계산해서 보여 준다. line 명령어로는 연도별 지출액 평균의 기본적인 선 그래프를 생성하였다. 데이터 브라우저를 통해 확인할 수 있는 화면과 산출되는 그래프는 아래 <그림 8>과 같다.

<그림 8> collapse 실행 후 자료 및 그래프



출처: 저자 작성

위와 같이 간단한 예시를 통해 collapse를 살펴보았다. 좀 더 현실적으로 코드를 작성하면 아래와 같이 연도별, 공여국별, 원조 유형별 지원액과 약정액의 합계를 축약하게끔 만들어 볼 수 있다. 먼저, 앞에서 메모리에 저장한 snapshot 1을 불러온다.

```
. snapshot restore 1
. collapse (sum) usd_disbursement usd_commitment, by(year donorname donorcode aid_t)
```

축약된 데이터지만 조건이 다소 복잡하므로, 데이터 역시 자료가 간단하지 않은 것을 확인할 수 있다. 다자기구를 제외하고 프로젝트(C01)의 총액만 국가별로 비교하기 위해 아래와 같이 donorcode가 900번대인 다자기구 삭제, C01 이외의 나머지 원조 유형은 모두 삭제한다. 마지막 명령어를 통해 미국과 영국의 지출액, 약정액에 대한 국가별 그래프를 그려 볼 수도 있다.

```
. drop if donorcode > 900
. keep if aid_t == "C01"
. line usd_disbursement usd_commitment year if donorname == "United States" | donorname == "United Kingdom", by(donorname)
```

현재 collapse를 통해 생성된 데이터는 panel long form으로, wide form으로 변경하기 위해서는 reshape를 사용한다. 편의상 약정액과 지출액의 변수 명을 간략하게 변경(rename)하였다.

```
. rename (usd_commitment usd_disbursement) (comt disb)
. reshape wide comt disb, i(donorname) j(year)
```

〈그림 9〉 reshape 실행 결과

(note: j = 2014 2015 2016)			
Data	long	->	wide
Number of obs.	141	->	53
Number of variables	6	->	9
j variable (3 values)	year	->	(dropped)
xij variables:			
	comt	->	comt2014 comt2015 comt2016
	disb	->	disb2014 disb2015 disb2016

출처: 저자 작성

〈그림 9〉를 살펴보면 기존에 있던 j변수 year가 약정액 및 지출액과 결합하여 comt2014~comt2016, disb2014~disb2016으로 새로운 변수가 생겼다. 자료의 형태도 살펴보면 공여국(donorname) 알파벳 순으로 지출액과 약정액이 연도별 가로 형(wide form)으로 변경되었다. 참고로, i변수에서 지정하지 않은 donorcode와 aid_t는 가장 뒤로 밀려났다. 최종적으로, 소수점과 자릿수 구분 등을 위해 아래와 같은 명령어를 수행한다.

```
. format %9.2fc disb2014-comt2016
```

format은 값이 보이는 형태를 변경하는 명령어로서 9는 총 자릿수를, 소수점 뒤의 2는 소수점의 자릿수를, f는 숫자형의 고정율, c는 천 자리마다 구분하는 쉼표를 넣는다는 의미다.

V. 결론 및 시사점

지금까지 개발협력 분야에서 빅데이터 활용을 위한 한국의 개발원조 데이터의 공개 현황, OECD 개발원조 데이터의 개괄적인 설명과 함께 Stata를 활용하여 OECD DAC의 개발원조 데이터를 효율적으로 활용하는 방안에 대해 간략하게 소개하였다. 최근 추세를 살펴보았을 때, 데이터 공개와 이를 활용하는 것에 대한 중요성이 높아지고 있으며, 개발협력 분야에서도 관련 문제를 해결하기 위해 데이터를 적극적으로 도입하기 시작하였다. OECD DAC의 개발원조 데이터의 경우에도 매년 약 20만 개 이상 추가되고 있으며, 현재까지 누적된 데이터 수만 해도 300만 개를 넘어가기 때문에 단순히 조회만 하거나 엑셀(Excel)을 사용하는 기존의 방법으로는 효율적인 데이터 처리가 불가능한 상황이 되었다.

따라서 엑셀(Excel)을 대체할 수 있는 통계 프로그램의 활용이 필요하게 되었고, 본 연구에서는 Stata를 통한 DAC 데이터의 활용 방법에 대해 논의하였다. 본 연구는 데이터 분석을 수행할 수 있는 다양한 분야에서 실용적이고 유용한(practical) 기여를 할 수 있다. 특히, 계량연구 분야에서는 본 연구에서 소개한 Stata의 몇 가지 명령어와 추가적인 작업을 통해 DAC 데이터를 효율적으로 분석할 수 있을 것이다. 또한 실무 분야에서는 Stata를 사용하여 개발원조 빅데이터에 접근해 자료를 얻은 후, 데이터를 축약(collapse)하고 또 변형(reshape)하여 최종적으로는 사용자가 만든 자료를 익숙한 엑셀(Excel)에서 다시 사용할 수 있으므로 현실적인 도움이 될 수 있다. 결국, 방대한 개발원조 데이터를 다루는 데 있어 필수적인 도구로는 Stata와 같이 대용량 데이터 처리가 가능한 통계 프로그램을 사용하는 것이 효과적이며, 이후 사용자의 필요에 따라 엑셀(Excel)을 보조적 수단으로 사용할 수도 있다.

본 연구의 시사점은 다음과 같다. 첫째, ODA 통계 활용과 관련된 분석 방법 논의 및 시도가 그동안 거의 없었다는 점에서 본 연구는 향후 ODA 통계의 효과적인 활용에 기여할 수 있을 것이다. 특히, ODA 데이터의 다양한 항목과 이들의 관계 및 의미는 실무 분야에서 관련 데이터를 자주 접하지 않는 사용자라면 정확히 이해하기 힘든 것이 사실이다. 따라서 본 연구에서는 OECD 개발원조 데이터에 대한 개괄적인 설명과 함께 주요 항목에 대해 자세한 설명을 제시함으로써 데이터 사용자들이 올바르게 ODA 통계자료를 사용하고 분석할 수 있도록 하는 지식 제고에 기여하였다.

둘째, Stata를 활용하여 ODA 통계자료를 다루면서 사용자들의 Stata 사용 방법과 ODA 데이터 분석의 실재를 알 수 있도록 하였다. 기본적인 명령어지만 개발협력 분야의 데이터를 예시로 활용함으로써 사용자들에게 실제 데이터가 어떻게 구축되어 공개되어 있는지를 보여

제I장

제II장

섹
터
포
커
스

제III장

주었다. 또한 Stata 예제 코드를 함께 제시하며 사용자들이 Stata를 쉽게 따라하고 이해할 수 있도록 하는 가이드라인 역할까지 하였다.

셋째, 향후 개발협력 데이터의 구축과 공개 방향 설정에 대한 필요성을 제시하였다. 선행연구들(김찬유, 2015; 김수진, 2016)과 함께 본 연구에서는 한국 개발원조 데이터의 공개 현황에 대해 간략히 살펴본 후 한계점을 제시하였다. 결국, 세계적인 데이터 공개 흐름에 맞추어 OECD 통계 홈페이지와 같이 원시자료 제공 서비스에 대한 수요는 더욱 늘어날 것이 분명해 보인다. 따라서 국내 개발원조 기관은 현재의 데이터 관리 수준의 강화를 통해 데이터의 품질 확보는 물론 데이터 공개로 이어질 수 있도록 노력해야 할 것이다.

다시 말해, 본 연구에서는 Stata를 이용하여 DAC 데이터를 다루는 방법에 대해 소개하였다. 그러나 최근 사용자가 늘어나고 있는 R이나 python과 같은 무료 프로그램으로 이러한 논의가 확장된다면 DAC 데이터를 필요로 하는 더욱 많은 사용자에게 도움이 될 수 있을 것이다. 그러므로 향후에는 효율적인 데이터 처리 방법(efficient method for data munging)에 대한 연구가 활발해지고, 관련 방법론을 논의하는 연구자가 생겨나길 희망한다.

〈참고 문헌〉

- 국무조정실. 2010. 『제7차 국제개발협력위원회 의결안건』(제7-2호~제7-5호). 세종: 국무조정실.
- 김수진. 2016. “SDGs 지표 데이터 이용가능성(Data Availability) 분석”. 『국제개발협력』 2016-2호: 79-113.
- 김찬유. 2015. “Post-2015와 지속가능한 개발을 위한 데이터 혁명”. 『국제개발협력』 2015-2호: 21-36.
- 민인식·최필선. 2015. 『STATA 기초적 이해와 활용』. 서울: (주)지필미디어.
- _____. 2016. 『STATA 패널 데이터 분석』. 서울: (주)지필미디어.
- 이인호. 2018 출간 예정. “Stata에서 World Bank Open Data의 효율적 활용 방안”. 『한국STATA학회』. 2018-1호.
- 조정화. 2018. “EDCF가 바라본 2017년 동료검토 주요 내용 및 향후 과제”. 『국제개발협력』 2018-1호: 21-34.
- 한국국제협력단. 2016. 『OECD DAC 통합통계보고지침』. 성남: 한국국제협력단.
- 황재상. 2018. “KOICA 차원에서의 2017 동료검토 주요 내용 및 향후 개선 과제”. 『국제개발협력』 2018-1호: 3-19.
- OECD. 2016a. “Converged Statistical Reporting Directives For The Creditor Reporting System (Crs) And The Annual Dac Questionnaire”, available at [https://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/DCDDAC\(2016\)3FINAL.pdf](https://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/DCDDAC(2016)3FINAL.pdf) (접속일: 2018.05.11.).
- _____. 2016b. “Converged Statistical Reporting Directives For The Creditor Reporting System (CRS) And The Annual Dac Questionnaire - Addendum 1”, available at [https://www.oecd.org/dac/stats/DCD-DAC\(2016\)3-ADD1-FINAL-ENG.pdf](https://www.oecd.org/dac/stats/DCD-DAC(2016)3-ADD1-FINAL-ENG.pdf) (접속일: 2018.05.11.).
- _____. 2017. “Data for Development: DAC Member Priorities and Challenges”, available at <https://www.oecd.org/dac/WP35%20Complete.pdf>(접속일: 2018.05.11.).
- KOICA 통계조회서비스: <http://stat.koica.go.kr/>(접속일: 2018.05.18.)

제I장

제II장

섹
터
포
커
스

제III장

ODA 통계조회시스템: https://stats.koreaexim.go.kr/index_outer.html(접속일: 2018.05.18.)

OECD Data 홈페이지: <https://data.oecd.org/>(접속일: 2018.05.18.)

OECD.Stat 홈페이지: <http://stats.oecd.org/>(접속일: 2018.05.18.)