

연구노트

학위논문에서 사용된 여론조사 자료의 품질평가에 관한 연구

A Study on the Evaluation for the Public Opinion Survey Data Quality Used in Theses

이해용* · 이인경**

Hae-yong Lee · In-kyeng Lee

본 논문은 학위논문에서 사용된 여론조사 자료의 품질을 평가하였으며, 또한 학위논문에서 사용되는 데이터의 품질평가 모형을 제시하였다. 데이터의 품질평가와 평가모형에 사용한 평가지표는 한국조사연구학회에서 제시한 조사윤리강령 제3조의 16가지 항목을 사용하였다. 연구결과 학위논문에서 사용된 여론조사 데이터의 품질은 낮은 수준임을 확인할 수 있었다. 16개 항목별로 석·박사학위에 따른 유의적인 차이가 있는지를 확인하기 위하여 두 집단 간 비율 차이 검정을 실시한 결과 표집오차와 분석방법의 수에서만 차이가 있을 뿐 다른 지표에서는 차이가 나타나지 않았다. 끝으로 학위논문에서 사용된 여론조사 자료의 질을 평가하기 위한 품질평가를 위해 9가지 주요 평가지표에 가중치를 부여하는 모형을 제시하였다.

주제어: 데이터품질, 조사윤리강령, 측정지표, 측정모형, 총족률

The purpose of this study is to evaluate the opinion survey data quality used in theses and to suggest the data quality evaluation model. We used 16 items which were suggested by KASR(Korean Association for Survey Research) as a standard of survey data quality evaluation in thesis. According to the result, the data quality used in the theses is very poor level not to be accepted. And we tried difference tests between two proportions in order to find there were significant differences among stochastic variables. The result showed that there were some significant differences in sampling error and number of analyzing method between doctor and master. we suggested a model that was weighted by 9 special items for a survey data quality evaluation in a thesis by observing them from KASR's principles.

* 교신저자(corresponding author): 성신여자대학교 통계학과 교수 이해용.

E-mail: hylee@sungshin.ac.kr

** 일진전기 전산사업부 기획팀

Key words : data quality, survey ethics principles, measuring items, data quality evaluation model, efficiency index

I. 서론

데이터가 재화가 되는 세상이다. 우리 주위에 널려 있는 데이터를 잘 정리하면 원하는 정보를 얻을 수 있고 그 정보를 활용하면 재화가 되기 때문이다. 그러나 정보라고 다 정보가 아니다. 또한 데이터라고 다 데이터가 아닌 것이다. 쓸모 있는 데이터를 이용하여 유용한 정보를 얻을 때 가능한 이야기다. “쓰레기를 넣으면 쓰레기가 나온다(garbage in garbage out)”는 말처럼 좋은 데이터를 넣어야 좋은 정보가 나오기 때문이다.

2000년대 들어 자료의 중요성이 대두되면서 조사기관을 중심으로 데이터 품질관리에 관심이 모아지기 시작하였다. 그동안 정확성과 신속성만을 강조했던 통계에서 적절성(relevance)과 전문성(professionalism), 그리고 공개성(openness) 등을 고려하여 조사의 모든 단계를 정보생산의 한 과정으로 보아야 한다는 인식이 확산된 것이다(류제복 외 2004). 국가통계는 물론이고 여론조사 자료에 대해서도 데이터 품질관리에 관심을 갖게 되었다. 최근에는 국가통계에 대한 품질관리가 실시되고 있으며, 조사업체에서도 자체적으로 조사 자료의 품질관리를 하고 있다.

공공기관뿐만 아니라 개인의 연구에도 자료의 중요성이 그 어느 때보다도 강조되고 있는 실정이다. 2005년도 줄기세포연구 논란이 사회적으로 커다란 파문을 일으키고 난 뒤에 자료의 중요성에 대한 세인들의 관심은 더욱 높아졌다. 개인의 연구논문에는 실험데이터와 조사데이터가 주로 사용된다. 일반적으로 개인의 연구논문에 사용되는 자료는 주로 논문 심사과정에서 검토될 것이라고 믿고 있다. 그러나 논문에 사용되는 데이터를 100% 신뢰하기에는 아직 그 환경이 좋지 못하다.

이에 본 논문에서는 여론조사를 실시하여 얻은 데이터를 이용하여 논문을 작성하는 사회과학 분야의 석·박사 학위논문을 대상으로 데이터품질의 현황을 알아보고 데이터의 품질을 평가하는 모형을 제시함으로써 데이터 품질관리연구의 기초를 제공하고자 한다.

II. 이론적 배경

“통계는 국가경영에 필수적인 인프라다”(손창균 2006)라는 인식이 팽배해지면서 데이터에 대한 관심이 그 어느 때보다 높다. 그러나 한편으로는 출처도 불분명한 데이터가 무분별하게 사용되고 있어 통계품질의 저하는 물론 사회적 불신마저 낳고 있다. 국가 대상 통계의 품질이 좋지 않으면 국가 경영에 차질을 빚게 되고, 일반회사의 경우에는 회사 경영에 막대한 손실을 가져올 수도 있다. 때문에 좋은 품질의 데이터를 만들어 이용하는 것은 국가 이익은 물론 개인의 이익과도 직결된다. 따라서 데이터 품질관리는 매우 중요하다. 늦은 감이 없지 않지만 최근 들어 각 국가나 기업 그리고 연구기관에서 데이터 품질관리의 필요성을 절감하고 품질관리에 대한 대책을 마련하고 있음은 매우 고무적이다.

데이터 품질관리가 중요하지만 품질을 떨어뜨리게 하는 요인은 도처에 도사리고 있다. 대표적인 예로 표본설계, 조사표설계, 무응답 대책, 조사원의 자질 등에서 오류가 발생할 수 있고, 보고통계에서는 보고자에게 유리하도록 왜곡된 데이터를 보고할 수도 있다(이기재 외 2006). 무엇보다도 조사환경이 날로 나빠지고 있는 가운데 수집되어야 하는 데이터의 질에 대한 기대치는 높아지고 있다. 따라서 신뢰할 수 있는 데이터를 생산하기 위해서는 데이터의 질을 관리할 수 있는 제도적인 보완과 더불어 국민의 의식이 성숙되어야 할 것이다. 그리고 객관적인 기준을 마련하여 계속적으로 관리하는 것이 바람직하다.

데이터품질의 관리는 1985년 캐나다에서 『통계품질 가이드라인』을 발간한 후 선진국 및 국제기관들이 관심을 보이면서 시작되었다. 그후 1995년 영국의 Bristol에서 “Survey Measurement and Process Quality”라는 주제로 열린 국제회의를 계기로 조사품질에 대한 인식이 크게 변하기 시작했다. 데이터의 품질에 대한 평가가 전통적으로 사용해 오던 평균 제곱오차(MSE) 중심의 평가방법을 벗어나, 데이터의 수집과정 전반에 걸쳐 관리하고 평가하는 것이 바람직하다는 인식을 갖게 되었다. 즉, 조사의 모든 단계를 정보생산의 한 과정으로 보아야 한다는 것이다(Collins & Sykes 1999).

그후 1998년 유럽의 통계청(EUROSTAT)장들이 “Quality Works and Quality Assurance within Statistics”라는 주제로 모임을 가졌다. 회의 결과 UN에서 승인하는 공식통계의 10대 기본원칙 외에 조사과정에서 데이터 질을 평가하는 데 중요한 요인으로 적절성과 전문성, 그리고 공개성을 강조하는 몇 가지 원칙이 포함되었다(Brackstone 1999). 각 기준은 대체적으로 차원(dimension) 또는 지표(indicator)와 요소(element)에 대한 개념을 중심으로 설정되어 있다. 다음 <표 1>은 각 나라와 국가 기관들이 작성한 데이터 품질평가지표 작성기준 현황을 보여주고 있다.

〈표 1〉 각국의 데이터 품질평가지표 현황

| 주 요인 | 한국 | 캐나다 | 호주 | 네덜란드 | IMF | Euro stat | OECD |
|----------|----|-----|----|------|-----|-----------|------|
| 정확성 | ○ | ○ | ○ | ○ | ○ | ○ | ○ |
| 시의성/정시성 | ○ | ○ | ○ | ○ | | ○ | ○ |
| 관련성 | ○ | ○ | ○ | ○ | | ○ | ○ |
| 접근성/편리성 | ○ | ○ | ○ | | ○ | ○ | ○ |
| 비교성 | ○ | | | | | ○ | |
| 효율성 | ○ | | | | | | |
| 일관성 | | ○ | | | | ○ | ○ |
| 해석가능성 | | ○ | | | | | ○ |
| 개편크기 최소화 | | | ○ | | | | |
| 커버리지 포괄성 | | | ○ | | | | |
| 비용효율성 | | | | ○ | | | |
| 응답부담 | | | | ○ | | | |
| 완결성 | | | | | | | |
| 신뢰성 | | | | | | | ○ |
| 방법론의 충실성 | | | | | ○ | | |
| 서비스성 | | | | | ○ | | |
| 무결성 | | | | | ○ | | |

* 출처: 이등명·김설희(2002)

또한 〈표 2〉는 여러 학자들에 의해 주장되어진 평가기준들을 열거한 것이다. 즉, 데이터가 신뢰받기 위하여 최소한 지켜야 할 사항들이다. 각국의 특성과 제반사항 등은 다르지만 일정한 평가기준이 지켜짐으로써 자료의 신뢰성이 확보될 수 있다는 것에 학자들의 본뜻이 있다고 할 수 있다.

최근 들어 우리나라에서도 조사자료에 대한 품질관리에 관심을 기울이고 있다. 통계청에서는 1999년 4월부터 통계품질관리업무를 시작하였으며, 2000년 12월에는 IMF와 공동으로 ‘통계품질평가 국제세미나’를 개최하여 정부통계의 품질평가 방안에 관한 자료를 수집하였다(Kim 외 2007). 여론조사에서 얻어지는 데이터 품질관리와 평가는 고전적인 데이터 품질관리기법으로 인식되어 오던 종래의 평균제곱오차(MSE)의 개념으로부터 데이터 작성 과정을 중심으로 하는 품질관리와 평가기법으로 바뀌고 있으며, 이런 현상은 선진국을 중심으로 자리매김하고 있다. 데이터의 질에 영향을 주는 주요지표는 〈표 1〉에서 제시한 바

〈표 2〉 학자들의 데이터 품질평가기준

| 학 자 | 평 가 준 거 |
|-------------------------|---|
| Ostrom(1979) | ①능률성(efficiency) ②효과성(effectiveness) ③형평성(equality) ④대응성(responsiveness) |
| Altman(1979) | ①능률성(efficiency) ②효과성(effectiveness) ③업무량(workload) |
| Brudney & England(1982) | ①경제적 측면 - 능률성(efficiency) - 효과성(effectiveness) ②정치적 측면 - 형평성(equality) - 대응성(responsiveness) |
| Morgan(1984) | ①능률성(efficiency) ②효과성(effectiveness) ③형평성(equality) ④대응성(responsiveness) |
| Rogers(1990) | ①경제성(economy) ②능률성(efficiency) ③효과성(effectiveness) ④서비스 수준(service level) ⑤대응(take up) |
| Amnons(1995) | ①업무량(workload) ②능률성(efficiency) ③효과성(effectiveness) ④생산성(productivity) |
| Eurostat(2000) | ①관련성(relevance) ②정확성(accuracy) ③시의성 및 정시성(timeliness & punctuality) ④접근성(accessibility & clarity) ⑤결합성(coherence) ⑥완전성(completeness) |
| IMF(2001) | ①통합성(integrity) ②정확성 및 신뢰성(accuracy & reliability) ③방법론적 견고성(methodological soundness) ④봉사성(serviceability) ⑤접근성(accessibility) |

• 출처: 이동명 · 김설희(2002)

있는 17가지 요인으로 요약될 수 있다. 그러나 통계에 문외한인 이용자들에게도 데이터 수집 과정에서 신뢰성을 유지할 수 있도록 권고하기 위해서는 세부사항에 대한 설명이 필요하다.

이러한 필요성을 인지하여 통계청은 통계품질평가지표를 8개의 절차별로 구분하고 세부 지표를 설명하여 통계이용자의 이해를 돕고 있다. 또한 ‘통계품질향상을 위한 평가지표의 개발’(류제복 외 2004)에서는 총 81개의 평가지표를 7개의 차원별로 분류하고 다시 각 차원은 몇 개의 부차원으로 세분화하여 상세한 평가지표를 제시하고 있다. 이동명·김설희(2002)는 ‘국가통계품질측정을 위한 체계적 접근의 품질평가’에서 8가지 중요단계에 각 단계별 48개의 주요지표를 정리한 바 있다.

이상에서 살펴본 바와 같이 국가통계의 품질관리에 대해서는 어느 정도 관심을 가지고 기초과정을 마련하였다고 볼 수 있다. 2000년에 들어서 한국조사연구학회에서 조사결과를 발표할 때 지켜야 할 ‘조사윤리강령’을 공포하였고(한국조사연구학회 홈페이지 www.kasr.org 참조) 그 후로 책임 있는 언론기관과 조사기관에서는 일부의 강령을 지키고 있는 추세다. 그러나 개인이나 단체 등에서 작성하여 연구에 사용하고 있는 데이터에 대해서는 관리 그 자체가 아예 없는 상태이므로 이에 대한 연구가 필요한 실정이다.

III. 학위논문에서 사용된 데이터의 품질 현황

학위논문에서 사용되고 있는 데이터의 품질이 어느 정도인가를 확인하기 위해서는 먼저 데이터의 품질을 평가하는 인증된 평가방법이 있어야 한다. 그러나 아쉽게도 인증된 평가방법이 없어서 본 연구에서는 조사연구학회에서 공포한 조사윤리강령 제3조의 16개 항목을 준수하고 있는지를 확인하여 평가방법으로 사용하였다. 여론조사 자료를 이용한 학위논문 145편을 연구대상으로 하고, 조사연구학회에서 공포한 조사윤리강령 제3조의 16개 항목을 평가지표로 사용하였다. 각 평가지표의 측정치는 16개 항목에 대한 내용이 학위논문 안에 언급되어 있는지 여부를 확인하는 방식으로 부여하였다.

1. 평가 설계

1) 조사 연구대상

본 연구대상 논문은 1990년대부터 2007년도에 걸쳐 작성된 국내 사회과학 석·박사학위 논문 중 국회도서관의 데이터베이스에 등재되어 있는 것으로 국한하였다. 구체적인 연구대

상 선정은 국회도서관 전자도서관 홈페이지에서 “소장 자료검색”, “학위논문”, “목차 여론조사”로 들어가 사회과학일반분야(DDC분류 300번) 가운데 연도별로 구분하여 검색한 결과 다음 <표 3>과 같은 총 160편의 논문을 연구대상으로 하였다.

2) 측정지표 선정 및 측정치 부여

데이터의 품질을 평가하기 위해서는 측정지표가 있어야 한다. 본 논문에서는 한국조사연구학회의 조사윤리강령에서 제시하고 있는 16개 지표를 이용했다. 조사윤리강령의 제3조에 의하면 “조사자는 조사방법이나 조사결과를 조사 의뢰자에게 보고하거나 일반인에게 공표할 때 다음 사항을 정확하게 밝힌다.” 라고 되어 있으며, 그 16개 항목은 <표 4>와 같이 기술되어 있다. 본 연구에서 이를 평가지표로 선정한 이유는 공신력 있는 기관에서 공표한 강령으로서 그 내용에 있어 객관적이고 타당성이 매우 높은 지표라고 판단되었기에 때문이다.

<표 3> 연구대상 논문의 내용

| 세부 사항 | 논문 수 | 세부사항 | 논문 수 |
|---|------|---|------|
| 학위년도: 2000~2007 학위구분: 석사 DDC분류: 사회과학(300) | 87건 | 학위년도: 2000~2007 학위구분: 박사 DDC분류: 사회과학(300) | 15건 |
| 학위년도: 1990~1999 학위구분: 석사 DDC분류: 사회과학(300) | 43건 | 학위년도: 1990~1999 학위구분: 박사 DDC분류: 사회과학(300) | 15건 |
| 합 계 | | | 160건 |

<표 4> 표본조사 결과 발표 시 공표해야 할 항목

| | | | |
|---|-------------|----|------------------|
| 1 | 조사자 | 9 | 조사방법 |
| 2 | 조사 의뢰자 | 10 | 질문지 |
| 3 | 조사목적 | 11 | 재통화, 재방문, 재발송 회수 |
| 4 | 조사시기 | 12 | 표본대체규칙 |
| 5 | 조사장소 | 13 | 응답률 |
| 6 | 모집단과 표집 틀 | 14 | 표집오차 |
| 7 | 표본크기 및 산정방법 | 15 | 가중치 부여방식 |
| 8 | 표집방법 | 16 | 기타 조사 및 분석절차 사항 |

• 출처: 한국조사연구학회, 조사윤리강령

측정지표가 선정되고 나면 측정지표에 값을 부여해야 한다. 어떤 지표에 어떤 값을 부여할 것인가를 결정하는 일이다. 예로 우리나라 통계청에서 현재 사용하고 있는 ‘통계 품질진단서(조사통계용)’를 보면 “매우 그렇다, 그렇다, 보통, 아니다, 전혀 아니다.” 등 5점 척도(관계없음 문항 제외)로 측정치를 부여하고 있다(통계품질진단서: 통계청 2008).

그러나 본 연구에서는 이상의 16개 지표에 대하여 학위논문 내에 언급이 되어 있으면 ‘1점’ 없으면 ‘0’점으로 간단히 부여하였다.

2. 측정결과 및 분석

본 연구는 사회과학의 석·박사논문(1990~2007) 160편을 조사한 결과 사용된 데이터가 여론조사가 아닌 것으로 판명된 15편을 제외한 145편을 대상으로 하였다. 16개 측정지표의 내용이 논문 안에 언급이 되어있는 경우 ‘1’점, 언급되어 있지 않은 경우 ‘0’점을 부여하여 16점 만점으로 측정하였다. 측정한 결과는 <표 5>와 같다.

<표 5>를 통하여 확인할 수 있는 것은 16가지 지표 중에서 조사자, 조사 의뢰자 및 조사 목적은 100% 논문에 기술되고 있다는 점이다. 학위논문은 열악한 조사환경에서 개인이 작성하는 것이므로 조사자와 조사 의뢰자는 동일인으로 간주하였기 때문이다.

이들 지표 외에 조사시기, 표본크기 및 산정방법, 질문지, 표본대체규칙, 응답률, 분석절차 등의 7가지는 50% 이상이 논문에 명기되어 있으나 재접촉회수, 표집오차, 가중치 부여방법은 단지 3% 미만이 언급되어 있는 실정이다. 이런 결과로 볼 때 현재 여론조사의 자료를 이용한 논문의 과반수 이상에서 데이터의 품질수준이 매우 낮다는 사실을 알 수 있다.

<표 5> 평가지표별 사용유무 빈도 표

| 평가지표 | 유 | 무 | 평가지표 | 유 | 무 |
|-------------|-----|-----|-----------|-----|-----|
| 조사자 | 145 | 0 | 조사 의뢰자 | 145 | 0 |
| 조사목적 | 145 | 0 | 조사시기 | 101 | 44 |
| 조사장소 | 50 | 95 | 모집단과 표집 틀 | 32 | 113 |
| 표본크기 및 산정방식 | 142 | 3 | 표집방법 | 53 | 92 |
| 조사방법 | 65 | 80 | 질문지 | 134 | 11 |
| 재접촉회수 | 2 | 143 | 표본대체규칙 | 116 | 29 |
| 응답률 | 124 | 21 | 표집오차 | 4 | 141 |
| 가중치 부여방식 | 3 | 142 | 분석절차 사항 | 94 | 51 |

〈표 6〉 학위별 평가지표 충족률

| 구분 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 총족률 |
|----|-----|-----|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 석사 | 1.0 | 1.0 | 1.0 | 0.70 | 0.35 | 0.20 | 0.98 | 0.33 | 0.45 | 0.92 | 0.01 | 0.81 | 0.86 | 0.01 | 0.02 | 0.60 | 0.58 |
| 박사 | 1.0 | 1.0 | 1.0 | 0.70 | 0.33 | 0.30 | 0.96 | 0.52 | 0.44 | 0.96 | 0.04 | 0.74 | 0.85 | 0.11 | 0.04 | 0.85 | 0.62 |
| 전체 | 1.0 | 1.0 | 1.0 | 0.70 | 0.34 | 0.22 | 0.98 | 0.37 | 0.49 | 0.92 | 0.01 | 0.80 | 0.86 | 0.03 | 0.02 | 0.65 | 0.53 |

〈표 7〉 학위별 평균 지표 비교

| 교차분석 | t-통계량 | p-value |
|--------------|---------|---------|
| 박사학위 vs 석사학위 | -2.0417 | 0.029 |

〈표 6〉은 석·박사 학위별로 16개 평가지표의 항목별 충족률(efficiency index)과 평균 충족률을 비교한 것이다. 〈표 6〉에서 석·박사 전체의 충족률은 석사가 58%, 박사가 62%, 전체 평균이 53%로 이를 100점 만점으로 환산해 보면 각각 58점, 62점, 53점 정도이다. 이것은 학위논문에서 사용하고 있는 여론조사의 데이터 질은 박사학위 논문은 60점을 조금 상회하고 있고 석사학위 논문은 60점 이하임을 나타내는 것이다.

이 결과를 이용하여 학위별로 유의적인 차이가 있는지를 확인하기 위하여 t-검정을 실시하였으며 그 결과는 〈표 7〉과 같다. 〈표 7〉을 보면 p-value가 0.029로 유의수준 0.05 하에서 박사와 석사는 차이를 보이고 있음을 알 수 있다.

다음으로는 석사와 박사 학위별로 각 16개 측정지표별로 충족률에 유의적인 차이가 있는지 알아보기 위해 두 비율 차에 대한 검정을 실시한 결과 〈표 8〉과 같다.

〈표 8〉에 의하면 16가지 지표 중 ‘표집오차’와 ‘분석방법의 수’에서만 5% 유의수준 하에서 석·박사 학위별로 차이가 있는 것으로 나타나고, 나머지 지표에서는 차이가 없는 것으로 나타났다.

‘표집오차’의 경우 박사학위 논문이 석사학위 논문보다 더 많이 이에 관해 기술하고 있었다. 또한 ‘분석방법의 수’를 보면 석사의 경우 분석방법의 기제가 전혀 없는 논문이 40% 정도를 차지하고 있다. 반면 박사의 경우 분석방법에 대한 기제가 없는 경우가 14%이고 그 외의 논문은 분석방법을 기제하였고, 분석방법의 수 또한 석사에 비해 많은 편이다. 물론, 분석방법 수가 많다고 좋은 것은 아니다. 하지만 최종후(1989)는 처리절차의 이름(procedure name)을 밝히는 것이 바람직하다고 권고하고 있다. 이 외의 지표에 있어서는

〈표 8〉 학위별 측정지표 차의 검정 결과표

| 평가지표 | Z-통계량 | p-value |
|-------------|-------|---------|
| 조사자 | 0.000 | 1.000 |
| 조사 의뢰자 | 0.000 | 1.000 |
| 조사목적 | 0.000 | 1.000 |
| 조사시기 | 0.090 | 04564 |
| 조사장소 | 0.140 | 0.556 |
| 모집단과 표집들 | 0.974 | 0165 |
| 표본크기 및 산정방법 | 0.525 | 0.700 |
| 표집방법 유무 | 1.783 | 0.037 |
| 조사방법 유무 | 0.044 | 0.518 |
| 질문지 | 1.072 | 0.142 |
| 재접촉회수 | 0.766 | 0.222 |
| 표본대체규칙 | 0.795 | 0.787 |
| 응답률 | 0.054 | 0.522 |
| 표집오차 | 1.681 | 0.046* |
| 가중치 부여방식 | 0.525 | 0.300 |
| 분석방법의 수 | 3.055 | 0.001* |

*는 유의수준 5% 하에 유의한 값

모두 유의수준 5% 하에서 각각의 임계치보다 작은 카이제곱 통계량을 갖게 되어 모두 유의하지 않기 때문에 석·박사 학위별로 차이가 없다고 할 수 있겠다.

이 외에도 표집방법이 명확하게 표기가 된 81부 중에서 유의추출(purposive sampling), 할당추출(quota sampling), 판단추출(judgement sampling)등 비확률 표본추출(non probability sampling)이 65부로 80%를 차지하고 있다. 이러한 추출방법은 조사원의 주관에 많은 영향을 끼치기 때문에 모집단의 특성에 대해 정확히 알고 있는 경우 등 제한적인 경우에 사용되어야 한다. 때문에 전문적 조사기관에서도 비표본오차 관리에 조심하는 추출법이다.

이상과 같이 조사연구학회의 윤리강령 16개 지표를 측정지표로 하여 학위논문에서 사용된 데이터의 질을 측정해 본 결과 석·박사 학위별로 약간의 차이는 있으나 전체적으로 볼 때 석·박사 논문 모두 사용된 데이터 품질이 의심스런 수준으로 판명되고 있다.

IV. 학위논문에서 사용된 데이터 질의 평가모형 제안

1. 평가지표의 선정

논문에서 사용된 데이터의 품질을 평가하기 위해서 평가지표를 선정하는 것은 단순하게 판단할 수 없는 일이다. 어떤 평가지표를 선정하느냐에 따라서 그 결과가 판이하게 달라질 수 있기 때문이다. 평가지표 선정에 많은 이론이 있을 수 있으나 본 연구에서는 조사윤리강령 제3조 16개 지표를 평가모형에서도 그대로 사용하는 것이 현재로서는 타당성이 있다고 생각하여 이를 평가지표로 선정하였다. 사회적으로 공신력 있는 기관에서 많은 연구와 논의를 거쳐 내놓은 윤리강령 16개 항목이 여론조사자료의 품질을 평가하는 데 최소한의 조건이면서 타당성이 있다고 생각하여 본 연구에서는 이를 평가지표로 선정한 것이다.

2. 가중치 부여방법

평가지표가 선정되면 다음으로는 각 지표에 대한 측정치의 부여방법이 마련되어야 한다. 각 지표에 어떤 기준하에, 어떻게, 얼마만큼의 값을 부여하느냐에 따라서 결과 또한 매우 민감하기 때문이다. 앞에서 평가지표로 선정한 16개 지표가 모두 동일한 중요도를 가지고 있다고 가정하면 이는 단순하다. 그러나 이렇게 가정하는 것은 무리가 있다. 예로 표본의 크기와 재접촉횟수의 중요도가 동일할 수는 없기 때문이다. 이점을 감안하여 학위논문에서 사용된 여론조사 자료에 대한 품질평가를 위해서 평가지표별 가중치를 부여하는 방법이 고려되어야 한다. 가중치 부여방법도 다양하며 또한 어떤 기준에 의하여 얼마의 가중치를 주느냐를 정해야 한다. 본 논문에서는 다음과 같은 방법으로 가중치를 부여하였다.

첫째, 윤리강령 16가지 평가지표 중에서 국내·외 여론조사 보도 시에 주로 공표하도록 하고 있는 ‘공직선거법 제 108조 5항’과 ‘해외마케팅 및 여론조사 수행규약 및 지침서’(이흥철 2005)에서 열거하고 있는 항목을 대상으로 조사자, 조사 시기, 모집단, 표본의 크기, 표집방법, 조사방법, 응답률, 질문지, 표집오차 등 9가지 지표를 주요지표로 선정하였다.

둘째, 주요 평가지표로 선정한 9가지 지표는 일률적으로 1.5배, 2배, 3배와 같이 가중치를 부여하고 나머지 7가지 지표는 가중치를 1로 동일하게 주는 방안을 고려하였다. 가중치를 고려하여 계산된 점수는 가중치에 따라 다른 값을 갖게 되므로 이해를 돕기 위하여 각각의 가중치를 10점 만점으로 환산하였다. 이와 같이 작성된 16개 평가지표별 가중치는 <표 9>와 같다.

〈표 9〉 각 평가지표별 가중치

| 평가지표 | 조정 전 | 1.5배 | 2배 | 3배 |
|--------------|-------|---------|-------|-------|
| 조사자* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 조사의뢰자 | 10/16 | 10/20.5 | 10/25 | 10/34 |
| 조사목적 | 10/16 | 10/20.5 | 10/25 | 10/34 |
| 조사시기* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 조사장소 | 10/16 | 10/20.5 | 10/25 | 10/34 |
| 모집단과 표집틀* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 표본크기 및 산정방법* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 표집방법 유무* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 조사방법 유무* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 질문지* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 재접촉회수 | 10/16 | 10/20.5 | 10/25 | 10/34 |
| 표본대체규칙 | 10/16 | 10/20.5 | 10/25 | 10/34 |
| 응답률* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 표집오차* | 10/16 | 15/20.5 | 20/25 | 30/34 |
| 가중치 부여방식 | 10/16 | 10/20.5 | 10/25 | 10/34 |
| 분석절차사항 | 10/16 | 10/20.5 | 10/25 | 10/34 |
| 합 계 | 10 | 10 | 10 | 10 |

*는 보도 기준에서 추천되는 중요 사항

또한 점수화의 목적은 설문 데이터의 신뢰성을 확인케 하는 데 있다. 때문에 논문 신뢰성의 우수여부를 판단할 수 있는 점수의 설계가 필요하다. 그 표현방식은 10점이나 100점과 같은 인식하기 쉬운 단위로 환산해 주거나 “수, 우, 미, 양, 가”와 같이 문자로 만들어 측정하는 방법이 있다. 따라서 본 연구에서도 두 가지 경우를 모두 검토하였다.

3. 데이터 품질평가모형

앞에서 언급한 내용(16개의 평가지표를 대상으로 9개 지표는 가중치를 부여하고 나머지 7개 지표는 가중치를 부여하지 않는 방법)을 기초로 하여 여론조사자료를 사용한 학위논문의 데이터 품질평가 점수 산출을 위한 모형을 만들면 다음의 식과 같이 설정할 수 있다.

$$\text{평가점수} = \sum_{i=1}^9 \alpha_i X_i + \sum_{j=1}^7 Y_j$$

여기서 α_i 는 상수(1.5, 2.0, 3.0)이고 사용된 변수는 아래와 같다.

| | | | |
|-------|-------------|-------|---------|
| X_1 | 조사자 | Y_1 | 조사의뢰자 |
| X_2 | 조사시기 | Y_2 | 조사목적 |
| X_3 | 모집단과 표집틀 | Y_3 | 조사장소 |
| X_4 | 표본크기 및 사정방법 | Y_4 | 재접촉횟수 |
| X_5 | 표집방법유무 | Y_5 | 표본대체규칙 |
| X_6 | 조사방법유무 | Y_6 | 가중치부여방식 |
| X_7 | 질문지 | Y_7 | 분석절차사항 |
| X_8 | 응답률 | | |
| X_9 | 표집오차 | | |

위 모형을 이용해서 가중치를 적용하지 않은 경우와 가중치를 1.5, 2.0, 3.0을 적용한 4가지 경우의 결과를 대상으로 10점 만점으로 환산하여 기초통계값을 구하였으며, 그 내용은 <표 10>과 같다.

<표 10>을 보면 가중치를 부여하지 않은 경우와 가중치를 부여한 후의 평균값들 사이에 약간의 차이가 있는 것으로 나타났다. 이들의 차이가 통계적으로 유의한 차이인지 아닌지를 확인하기 위하여 SPSS를 이용하여 분산분석을 하였으며 그 결과는 <표 11>과 같다. 검정결과 4가지 집단간의 차이는 없는 것으로 판명되었다.

<표 10> 가중치를 부여한 후의 통계(10점 만점으로 환산한 경우)

| | 연구대상 논문수 | 최소값 | 최대값 | 평균 | 표준편차 |
|------------|-------------|------|------|--------|---------|
| 가중치 없음 | 145 | 2.50 | 8.13 | 5.8491 | 1.23839 |
| 주요지표에 1.5배 | 145 | 2.20 | 8.29 | 5.9108 | 1.28065 |
| 주요지표에 2.0배 | 145 | 2.00 | 8.40 | 5.9503 | 1.31562 |
| 주요지표에 3.0배 | 145 | 1.76 | 8.53 | 5.9980 | 1.36522 |

<표 11> 분산분석 결과

| | 제곱합 | 자유도(df) | 평균제곱 | 유의확률 |
|------|---------|---------|-------|--------|
| 집단 간 | 1.726 | 3 | 0.575 | 0.7961 |
| 집단 내 | 974.644 | 576 | 1.692 | |
| 합 계 | 976.370 | 579 | | |

〈표 12〉 145개 조사대상 논문의 데이터 품질평가 결과

| | 수 | 우 | 미 | 양 | 가 | 계 |
|------|---|----|----|----|---|-----|
| 조정 전 | 0 | 20 | 80 | 44 | 1 | 145 |
| 1.5배 | 0 | 33 | 79 | 32 | 1 | 145 |
| 2.0배 | 0 | 33 | 79 | 32 | 1 | 145 |
| 3.0배 | 0 | 38 | 71 | 35 | 1 | 145 |

또한 평가의 결과를 “수, 우, 미, 양, 가”의 5가지 등급으로 표현해 보았다. 가중치를 고려한 3가지와 고려하지 않은 경우를 대상으로 10점 만점으로 환산된 점수에서 0점에서 2점을 ‘가’, 3점에서 4점을 ‘양’, 5점에서 6점을 ‘미’, 7점에서 8점을 ‘우’, 9점에서 10점을 ‘수’로 평가하였는데 그 결과는 〈표 12〉와 같이 나타났다.

〈표 12〉를 보면, 가중치를 부여하게 되면 전체적으로 평가점수가 ‘양’에서 ‘우’로 바뀌는 것을 볼 수 있다. 즉 논문에 사용된 데이터의 질을 평가해 볼 때 가중치를 3배 하게 되면 조정하지 않은 경우나 1.5배나 2.0배의 경우에 비해 ‘우’가 증가하는 경향을 보이고 있다. 어느 경우도 ‘수’에 해당하는 논문이 없다. 이러한 이유는 조사논문의 질이 기본적으로 좋지 않았다는 의미이다. 다만 주요 평가지표에 가중치를 부여하게 되는 경우 논문에 사용된 데이터의 질적 평가가 조정하지 않은 경우에 비하여 약간 향상되는 효과는 나타나고 있다.

V. 결론 및 향후 연구방향

본 연구의 결과는 크게 네 가지로 요약할 수 있다. 하나는 여론조사 자료를 이용한 기존의 사회과학 논문들을 대상으로 데이터 품질을 평가해 보는 것으로, 분석 결과 예상했던 것 이상으로 석·박사 구별없이 데이터 품질수준이 낮다는 것을 확인했다. 둘째는 학위논문에서 사용되는 여론조사 데이터의 품질을 높이기 위한 방안으로 학위논문에서 사용된 데이터 품질을 객관적으로 평가하는 것이 필요하다는 판단에 따라 측정모형을 제시하였다. 셋째는, 평가모형서 평가지표에 가중치를 부여하는 경우와 그렇지 않은 경우를 고려한 예로 연구대상으로 선정한 논문 145편을 대상으로 평가하는 방법과 그 평가결과를 제시하였다. 주요 평가지표에 가중치를 부여한 결과 그렇지 않은 결과와 크게 차이가 없음을 확인하였다. 끝으로 한국조사연구학회의 조사윤리강령에서 제시한 16가지 지표를 준거로 학위논문에서 사용된 여론조사 데이터의 품질을 평가해 보려는 시도를 하였다.

본 연구는 한국조사연구학회 조사연구윤리강령 제3조의 16개 항목을 평가지표로 하고 평가지표의 측정은 논문에서 각 평가지표에 대한 명기 여부를 기준으로 간단하게 평가함을 기준으로 하였기 때문에, 다른 평가지표와 다른 측정방법을 사용할 경우 얼마든지 다른 결과를 얻게 될 수 있다는 한계점을 가지고 있다. 본 연구에서 사용한 16개 평가지표가 절대적이지 않고 또한 어떤 과학적인 근거에 의해 선정된 것이 아니기에 다른 지표를 얼마든지 가감할 수 있는 여지가 있다. 평가지표의 중요도가 각각 다른 상태에서 어느 정도는 임의성을 가지고 가중치를 부여한 점과 그를 기초로 평가모형을 만들어 제시한 점에서 논란의 여지가 있다. 16개의 평가지표에는 표본의 크기, 재통화횟수, 응답률, 표본오차 등과 같이 양적으로 나타낼 수 있는 데이터와 조사자나 조사 의뢰자와 같이 이름을 나타내는 데이터 등으로 구성되어 있다. 이런 데이터의 특성을 고려하여 가중치를 부여하는 방안을 더 연구할 필요가 있다.

더 많은 논문을 조사하여야 함에도 불구하고 한편의 논문을 대상으로 16가지 평가항목에 대하여 언급되어 있는지를 일일이 확인하여 자료를 구축하는 과정이 쉽지 않았다. 145개의 논문을 정독하여 관련 사실을 확인해야 하는 제약으로 인해 전 학문분야가 아닌 사회과학분야에만 국한하여 평가를 진행한 것이 아쉬운 점이다. 때문에 향후연구에서는 타 분야는 물론 연속변수의 세부적인 단계를 제시하고, 조사방법별로 판단점수를 달리 제시하는 등의 새로운 아이디어 창출이 있기를 바란다.

참고문헌

- 류제복·유정빈·김선우. 2004. “통계품질향상을 위한 평가지표의 개발.” 《조사연구》 5(2): 85-107.
- 손창균. 2006. “한국복지패널에서 응답형태에 따른 패널가구의 특성 비교 연구.” 《한국복지패널 심층분석 보고서 2006》.
- 이기재·김영원·박진우·김설희·박은영. 2006. “보고통계품질향상을 위한 평가 지표의 개발.” 《조사연구》 7(1): 85-107.
- 이동명·김설희. 2002. “국가통계품질측정을 위한 체계적 접근의 품품질평가.” 《한국조사연구학회 추계학술대회 발표논문집》 111-127.
- 이홍철. 2005. 《해외마케팅 및 여론조사 수행규약 및 지침서》. 한국마케팅여론조사협회, 엔아이코리아 표준화 연구팀.
- 최중후. 1989. “국내 과학학술지의 통계적 기법 활용에 관한 신뢰성 평가.” 고려대학교 박사학위 논문.

통계청. 2008. 《통계품질진단서》. 통계청.

Brackstone, G. 1999. "Managing Data Quality in a Statistical Agency." *Survey Methodology* 25(2): 139-149.

Collins, M. and W. Sykes, 1999. "Extending the Definition of Survey Quality." *Journal of Official Statistics* 15(1): 57-66

Kim, Sootack, Kiho Jeong, and Seolhee Kim, 2007. "Developing the Quality Assessment Indicators for the National Processing Statistics of Korea." *The Korean Communications in Statistics* 14(3): 649-665.

[접수 2010/5/20, 수정 2010/7/12, 게재확정 2010/7/15]