

국가승인통계 품질 등급 부여 및 상대지표 개발

심규호[†]

* 통계개발원 동향분석실

Research on grading the quality level and developing the comparability index of the national statistics

Kyu Ho Shim[†]

* Trend Analysis Division, Statistical Research Institutes, Statistics Korea

Key Words : statistics, national statistics, quality, quality dimension, quality level

Abstract

Statistics Korea has been diagnosis national statistics every year since 2006. They diagnosis over 200 kinds of national statistics. They have 7 quality dimension used for quality diagnosis. That is relevance, accuracy, Timeliness, Comparability, Coherence, Accessibility. Since we are interest in how well they produce national statistics, comparability has become the most important dimension recently. In this reason, Statistics Korea try to rating quality level and development comparability index for national statistics. This paper propose the practical method of grading the quality level and developing the comparability index of the national statistics

1. 서 론

최근에 국가 차원에서의 통계에 대한 인식이 높아지고 통계청뿐만 아니라 많은 국가기관에서 통계 사용자의 요구를 만족시키기 위한 통계 개발에 박차를 가하고 있다. 통계청은 중앙 통계기관으로서 이러한 통계 개발을 선도하고 관리해 나가려고 노력하고 있다. 이에 따라 2006년에 국가승인통계 106종 품질진단을 실시를 시작으로 2007년에는 180종, 2008년에는 170종 2009년에는 58종의 국가승인통계에 대한 품질진단을 실시하고 있다. 이를 실시하기 위해 매 실시 년도에 각 해당 통계의 전문가로 구성된 품질 진단 전문가 팀이 구성된다. 이 품질 진단 팀은 각 분야의 통계의 정의, 생산과정, 사용자들의 의견, 제공 통계의 정확성, 정책에서의 사용 정도 등 많은 통계 품질과 관련된 정보를 취합하고 분석하여 품질 개선에 도움이 되는 보고서를 작성하고 있다. 이러한 보고서에는 품질이 좋은 부분은 우수

사례로 선정하여 공표하기도 하고 개선해야 할 점은 보고서에서 상세한 개선 방법을 제시하기도 한다. 이 보고서를 기반으로 통계 생산 기관이나 부서에서는 품질을 개선하여 국가통계 전반에 기여하기를 기대하는 것이다.

품질 관리 담당자들이 느끼는 어려움은 다음의 두 가지이다. 먼저 품질 개선 활동은 열심히 하고 있지만 품질 개선에 대한 명확한 평가를 하기가 어렵다는 것이다. 개선 방법에 대한 것은 보고서에 많이 기재하여 통계 생산 부서에 전달하지만 그 내용처럼 실제적으로 개선을 강제할 법적 권한은 없다. 대부분의 통계 생산 기관은 통계 인력 또는 통계 예산 측면에서 열악한 환경을 가지고 있는 경우가 많다. 또한 품질 진단 자체가 어떠한 비교할 수 있는 점수화나 비교 지수를 개발하여 공표하여야 하는데 현실적인 어려움이 있다. 우선 품질을 정확하게 판단하고 개선하는 데는 두 가지가 선행되어야 한다. 첫째는 품질의 각 성격을 대표하는 지표가 측정 가능해야 한다. 유럽에서는 이미 품질을 측정하는 방법에 대한 연구가 많이 진행되었다. 통계 생산에서부터 공표, 활용 부분 전반에 걸쳐 상세한 측정 기준에 따

[†] 교신저자 khshim@korea.kr

라 측정하고 점수화를 한다. 또한 이러한 점수화된 결과를 분석하여 품질에 영향을 많이 미치는 지표를 찾아 내기도 하였다. 둘째는 이 측정된 지표들이 비교 가능해야 한다. 또한 정확한 측정된 값으로 비교 가능한 종합 지표를 개발하여 진단된 통계들 간의 서로 비교가 가능해야 한다. 서로 비교 가능한 지표들이 개발되면 통계 생산 주체들 간에도 비교와 분석을 통해 품질이 개선되는 효과가 있으리라 기대한다.

본 연구에서는 품질관리의 현황을 국내와 국외의 현황으로 나누어 비교 분석하고자 한다. 국내의 품질 관리는 어떻게 진행하고 있는지 문제점은 없는지 알아보고 해외의 품질 관리 현황과 연구 현황에 대해 알아보고자 한다. 국내 사례는 통계청 품질 진단을 중심으로 알아보고자 한다. 해외에서는 EUROSTAT과 캐나다 등이 품질진단의 원조 격이라 할 수 있고 많은 품질 진단에 대한 연구가 진행되어져 왔다. 해외의 품질 진단 현황에 대해서는 유럽, 영국, 이탈리아를 중심으로 알아보고자 한다.

2. 통계청 품질 진단

통계청의 품질관리는 품질관리과에서 맡고 있다. 1999년 4월 통계기획국 기획과에 품질평가팀이 신설되었고 2002년 7월 청장직속 부서인 품질관리팀으로 개편, 2005년 품질관리팀에서 품질관리과로 확대 개편되었다. 2006년에 국가승인통계 106종 품질진단을 실시를 시작으로 2007년에는 180종, 2008년에는 170종 2009년에는 58종의 국가승인통계에 대한 품질진단을 실시하고 있다.

통계청에서는 통계품질 차원을 다음의 6개 차원으로 정의하고 있다. 통계청에서는 관련성, 정확성, 시의성, 비교성, 일관성, 접근성을 차원으로 정하였으나 국가마다 조금의 차이가 있다.

이러한 품질 차원은 품질 진단을 하는데 있어 중요한 요소가 되며 이러한 차원을 바탕으로 품질 진단을 실시하며 각 차원 별로 개선 사항을 도출하여 결과적으로 품질 향상에 기여하고 있다.

3. 해외 측정 가능 품질 지표 연구

3.1 EUROSTAT

EUROSTAT이 추진한 연구의 목적은 European

<표 1> 통계청 품질 차원

차원	설명
관련성 (Relevance)	이용자의 요구사항을 파악하여 통계자료에 반영하는 정도
정확성 (Accuracy)	측정하고자 하는 모집단의 특성이나 크기를 얼마나 정확하게 측정했는가.
시의성 (Timeliness) 및 정시성 (Punctuality)	시의성은 통계작성기준시점과 결과발표시점 간의 차이에 대한 개념 정시성은 예고된 공표일정을 준수하는가에 대한 개념
비교성 (Comparability)	시간적·지리적 또는 포괄하는 영역이 달라도 어떠한 통계가 개념, 분류, 측정도구, 기초자료 등이 동일하게 적용되어 개념적으로 서로 비교가 가능한지를 측정
일관성 (coherence)	일관성은 동일한 경제·사회현상에 대해 서로 다른 기초자료나 작성방법에 의해 작성된 다른 통계 또는 소관통계가 얼마나 유사한가를 측정
접근성 (accessibility) 및 명확성 (clarity)	접근성은 이용자가 통계자료에 얼마나 쉽게 접근할 수 있는지에 대한 정도 명확성은 통계가 어떻게 만들어졌는지에 대한 정보수준

Statistical System에서 시간에 따라 생산되는 자료의 품질을 측정할 수 있는 제한된 지표의 집합을 가지는데 있다. 그 지표는 EUROSTAT의 품질 정의에 정의된 품질 차원을 표현하여야 하며 또한 그 지표들을 계산하는 방법론은 잘 구축되어야 하며, 쉽게 설명할 수 있어야 한다. 지표를 정의할 때에는 기준(criteria)을 준수하지 않은 품질에 대한 결과물의 차원에 대해서도 측정가능하게 하기 위해서 모든 지표들이 시험적으로 고려되어야 된다는 것을 알아냈다. 따라서 지표들은 Key Indicators, Supportive Indicators, Indicators for further Indicator로 분류되어야 한다. Key Indicators는 조건(criteria)을 무조건 만족시켜야 할 의무가 있다. Supportive Indicators는 데이터 품질을 간접적으로 측정하기 때문에 중요하게 여겨지는 만큼 조건을 만족시켜야 한다. 나머지 세 번째 Indicators들은 심층적인 경험과 조사기관들 간의 토론에 목적이 있다. 서로 다른 주제 영역에 대한 통계의 품질을 모니터링하기 위한 스코어보드를 정의할 때는 지표의 목록이 고려되어야 하며, 지표들이 연관성이 있어야 하고 특정 영역이 선택되도록 구현되어야 한다. 서로 다른 품질의 영역의 올바른 균형을 유지하고 최소한의 지표수를 유지하기 위해 세심한

관심이 요구된다. 이러한 메타데이터의 정보는 품질 보고서에 항상 담겨 있어야 한다. 사용자들에게 통계의 품질에 대해 알려주기 위해 지표를 사용할 때에는 Qualitative statements, 품질 정보를 설명하는데 필요한 지원, 통계 사용의 주요 효과에 대한 요약을 포함할 것이 요구된다. (EUROSTAT, 2005)

<표 2> 품질 영역과 지표

Quality Component	Indicator	Category
Relevance	R1. User satisfaction index	3
	R2. Rate of available statistics	1
Accuracy	A1. Coefficient of variation	1
	A2. Unit response rate (un-weighted/ weighted)	2
	A3. Item response rate (un-weighted/ weighted)	2
	A4. Imputation rate and ratio	2
	A5. Over-coverage and misclassification rates	2
	A6. Geographical under-coverage ratio	1
	A7. Average size of revisions	1
Timeliness and Punctuality	T1. Punctuality of time schedule of effective publication	1
	T2. Time lag between the end of reference period and the date of first results	1
	T3. Time lag between the end of reference period and the date of the final results	1
Accessibility and clarity	AC1. Number of publications disseminated and/ or sold	1
	AC2. Number of accesses to databases	1
	AC3. Rate of completeness of meta data information for released statistics.	3
Comparability	C1. Length of comparable time-series	1
	C2. Number of comparable time-series	1
	C3. Rate of differences in concepts and measurement from European norms	3
	C4. Asymmetries for statistics mirror flows	1
Coherence	CH1. Rate of statistics that satisfies the requirements for the main secondary use	3

* Category : 1 = Key, 2 = Supportive, 3 = Further Experience

제공되는 통계적인 정보에 대한 포괄적인 품질을 평가하기 위한 한 개의 합성 지표를 개발하는 것에 대한 다양한 의견이 존재한다. 이러한 지표가 개발된다면 각각의 기관에 제공되는 통계의 종합적인 품질을 평가하는데 도움이 되며 모든 기관에 제공되는 종합적인 품

질 정보를 비교하기 위해 결집하는데 도움이 된다. 기술적인 면에서 보면, 지표가 어떤 한 나라의 포괄적인 품질을 평가하는 것은 가능해 보이며 심지어는 각각의 관계되는 가중 품질 지표를 개발함으로써 쉬워 보이기도 한다. 그러나 개념적이고 실제적인 관점에서 보면 EU와 같이 산재된 곳의 포괄적인 통계 품질을 평가하는 나라간의 많은 차이점이 존재하고 모든 조사 단계에 있어 미리 정의된 표준이라는 것이 존재하지 않기 때문이다. 또한 어떤 표준이 존재하면 항상 따르지 않기 때문이다. 그러므로 각국은 그들의 요구에 의존된 각각의 품질 요소를 평가하기 위한 서로 다른 가중치를 적용해야 한다. 그리고 물론 각각의 요소의 평가에는 비용과 부담이 요구된다.

3.2 Office for National Statistics (영국통계청)

Guidelines for measuring statistical quality(version 3.1)는 통계적인 산출물의 품질을 측정하거나 보고할 때 사용되는 품질 측정값들과 지표의 체크리스트를 제공한다. 이것들은 국가통계(National Statistics)는 아니며 통계적인 산출물의 생산 전반에 품질을 측정하는데 가장 좋은 방법을 표현하고 있다. 이 새로운 버전에는 비밀보호 기법이 추가됨으로 인한 정보 손실에 대한 측정값이 새로 추가되었다. 이 문서의 목적은 Government Statistical Service(GSS) 전반에 걸쳐 품질을 측정하거나 보고하는데 필요한 표준화된 접근법을 제안하는데 있다. (Karen, 2007)

GSS는 사용자들에게 통계적인 산출물을 수집하고 편집하는데 사용되는 기법에 대한 정보를 제공할 것을 위임 받았다. 이러한 위임에 대해서는 National Statistics code of practice¹⁾에 표현되어 있다.

‘National Statistics를 생산하는 과정과 기법은 사용자가 각각의 목적에 적합하게 접근할 수 있을 정도로 충분히 세밀해야 한다.’

여기에 더해 GSS는 National Statistics Quality Strategy에 자신들의 목적이 통계적 산출물의 품질에 대한 정보를 제공하는 것임을 공표하였다. ‘National Statistics에 대한 품질의 측정은 시스템적으로 보고되어야

1) 영국통계청에서 반드시 따라야 하는 주요 이론과 표준을 제시한 문서
(http://www.statistics.gov.uk/about_ns/cop/downloads/professionalcompetence.pdf)

하며, 사용자가 사용하고자 하는 그들의 분야에서 알맞은 것인지에 대한 평가를 할 수 있어야 한다. 아래 <표 3>은 문서에 포함된 품질 차원이다. 품질 차원 별로 측정되는 값들이 잘 정리되어 있다.

<표 3> 품질 차원

Definition	Key components
1. RELEVANCE	
The degree to which the statistical product meets user needs for both coverage and content.	Any assessment of relevance needs to consider: <ul style="list-style-type: none"> • who are the users of the statistics; • what are their needs; and • how well does the output meet these needs?
2. ACCURACY	
The closeness between an estimated result and the (unknown) true value.	Accuracy can be split into sampling error and non-sampling error, where non-sampling error includes: <ul style="list-style-type: none"> • coverage error; • non-response error; • measurement error; • processing error; and • model assumption error.
3. TIMELINESS AND PUNCTUALITY	
Timeliness refers to the lapse of time between publication and the period to which the data refer. Punctuality refers to the time lag between the actual and planned dates of publication.	An assessment of timeliness and punctuality should consider the following: <ul style="list-style-type: none"> • production time; • frequency of release; and • punctuality of release.
4. ACCESSIBILITY AND CLARITY	
Accessibility is the ease with which users are able to access the data. It also relates to the format(s) in which the data are available and the availability of supporting information.	Specific areas where accessibility and clarity may be addressed include: <ul style="list-style-type: none"> • needs of analysts; • assistance to locate information; • clarity; and • dissemination.
5. COMPARABILITY	
The degree to which data can be compared over time and domain	Comparability should be addressed in terms of comparability over: <ul style="list-style-type: none"> • time; • spatial domains (e.g. sub-national, national, international); and • domain or sub-population (e.g. industrial sector, household type).
6. COHERENCE	
The degree to which data that are derived from different sources or methods, but which refer to the same phenomenon, are similar.	Coherence should be addressed in terms of coherence between: <ul style="list-style-type: none"> • data produced at different frequencies; • other statistics in the same socio-economic domain; and • sources and outputs.

<표 4>의 주요 품질 측정값(KQMs : Key Quality

Measures)는 사용자에게 전반적인 품질 결과에 대한 가장 중요한 정보를 제공해주는 지표들에 대한 품질 측정값이다. 더 나아가 KQMs는 품질에 대한 성능을 모니터링하고 통계적 결과물에 대한 어떠한 품질 개선에 대한 관리 정보를 제공하여 준다.

<표 4> 주요 품질 측정값들

	KEY QUALITY MEASURE	ESS QUALITY DIMENSION	GUIDELINES REFERENCE
1	Where possible, describe how the data relate to the needs of users	Relevance	B1.3
2	Provide a statement of the nationally/internationally agreed definitions and standards used	Comparability	B1.13
3	Unit response rates by sub-groups, weighted and unweighted	Accuracy	B3.4 (Household surveys) B3.5 (Business surveys)
4	Key item response rates	Accuracy	B3.7
5	Total contribution to key estimates from imputed values	Accuracy	B4.7
6	Editing rate (for key items)	Accuracy	B4.11
7	Estimated standard error for key estimates	Accuracy	B5.2 (for key estimates of level) B5.3 (for key estimates of change)
8a	Time lag from the reference date/period to the release of the provisional output	Timeliness and Punctuality	B8.1
8b	Time lag from the reference date/period to the release of the final output	Timeliness and Punctuality	B8.2
9	Estimated mean absolute revision between provisional and final statistics	Accuracy	B8.21
10	Compare estimates with other estimates on the same theme	Coherence	B8.28
11	Identify known gaps between key user needs, in terms of coverage and detail, and current data	Relevance	B8.29

<그림 1>은 정확성에 대한 주요 품질 측정값의 한 예시이다. 사업체 조사에서의 하위 그룹별 응답률에 대한 측정값이다. 각각의 응답률에 대한 비율을 계산하여 가중 값을 계산하고 있다. 이 지표는 무응답 오차가 얼마나 유의한가를 보여주는 지표가 된다. 정교한 계산법

으로 품질을 측정하고 있음을 알 수 있다.

3.3 이탈리아 통계청

유럽과 NSI수준에서의 질적 측정에 대한 품질조사 품질이 평가 되어 왔다. 개념적으로는 이러한 질적 측정이 EUROSTAT의 품질component에 관련이 되어 있지만 지표들 간의 품질component간에는 연관성은 완전히 조사되지는 않았다. 이 연구에서는 표준품질 지표들이 품질Component와 조사품질 전반적인 것을 잘 반영하고 있는지를 연구하고자 하였다.(Giovanna, 2008) 구조 방정식 모형(SEM : Structural Equation Modeling)은 품질 지표와 같이 관측된 변수와 품질 또는 품질 성분과 같은 관측되지 않은 변수사이의 관계를 분석하는 방법론적 접근 방법이다. 여기서 품질 또는 품질 성분은 이론적으로 직접적으로 측정 불가능한 요인을 말한다. 따라서 품질 지표와 같은 변수를 통해 간접적으로 관측하여야 한다. 이 연구에서 품질은 하나 또는 하나 이상의 표준 품질 지표의 수준에 영향을 받는 것으로 표현됨을 가정하고 있다. 다른 말로 품질은 하나 또는 그 이상의 품질 지표들 간의 공분산으로부터 나온 측정되지 않은 품질 성분에 의해 표현됨을 가정한다.

<그림 2>는 SEM 분석 결과를 나타낸다. 위의 분석 결과로서 Relevance와 Accuracy의 품질 관계가 어떻게 이루어지는지 알 수 있다. 분석 결과로서 우리는 품질 지표들이 표현하기 굉장히 복잡함에도 불구하고 품질 영역을 잘 표현하고 있음을 알 수 있다.

4. 통계청 품질 진단 분석

통계청에서 하고 있는 품질관리에 대한 분석을 하기 위해서는 기본적으로 통계청 승인 통계 품질 관리의 절차와 프레임에 대해 알아볼 필요가 있다. 품질관리과에서는 품질관리기반, 이용자 만족도 및 요구사항 반영 실태, 세부 작성절차별 체계, 수집 자료의 정확성, 통계 자료서비스의 충실성의 5가지 단계를 통해 관리를 하고 있다. 각 분야의 전문가를 통해 이러한 작성 단계별로 정보를 수집하여 분석하고 개선 사항을 만들어 반영하여 보고서를 작성하는 것을 목표로 한다. 이러한 단계별 작성 과정을 분석하여 품질에 영향을 미치는 요인들이 정확하게 측정되고 있는지 품질 평가에 있어 누락된 요인들은 없는지를 찾아서 개선해 보고자 한다. 이 장에서는 전반적인 품질 관리의 단계별 내용과 성격에 대

해 알아보고 다음 장에서는 실제 평가되고 보고된 보고서를 분석하여 세부 개선사항에 대해 중점적으로 알아보고자 한다. 분석 방법은 통계청에서 진단하여 보고하고 있는 품질진단 보고서를 분석하였다. 분석 대상은 2008년 10월에 발간한 “2008년 정기 통계품질진단 연구용역 최종 결과보고서”이며 영역은 「기업결합동향」, 「문화재관리현황」, 「재래시장실태조사」, 「농림업 생산지수」의 네 개 영역이다. 세부적인 보고서의 차이점을 살펴보면 기업결합동향과 농림업 생산지수에는 해외사례가 있으나 다른 두 보고서에는 없다.(통계청 품질관리과, 2008) 해외사례라는 것이 품질 진단 영역마다 있는 경우도 있고 없는 경우도 있을 것이지만 보고서의 큰 흐름에 있어서 언급조차 없다는 것은 의아한 부분이다. 또한 품질에 있어서 데이터의 품질이 가장 중요하다고 판단되지만 보고서는 데이터의 품질에 대한 부분보다는 다른 이용자의 측면이나 정책반영 부분 등과 같은 데이터 자체의 품질 외적인 부분에 더 진단이 치중되어 있음을 알 수 있다. 품질 진단 보고서를 분석하는 목적은 첫째, 품질 진단 단계별 사항을 확인하여 서술 식으로 되어 있는 보고서 진단 결과를 최대한 수식 화하여 보고서별 비교가 가능하게 하려고 한다. 두 번째는 보고서 전반에 걸쳐 세부적인 사항에 대해 개선할 점을 제시하고자 한다.

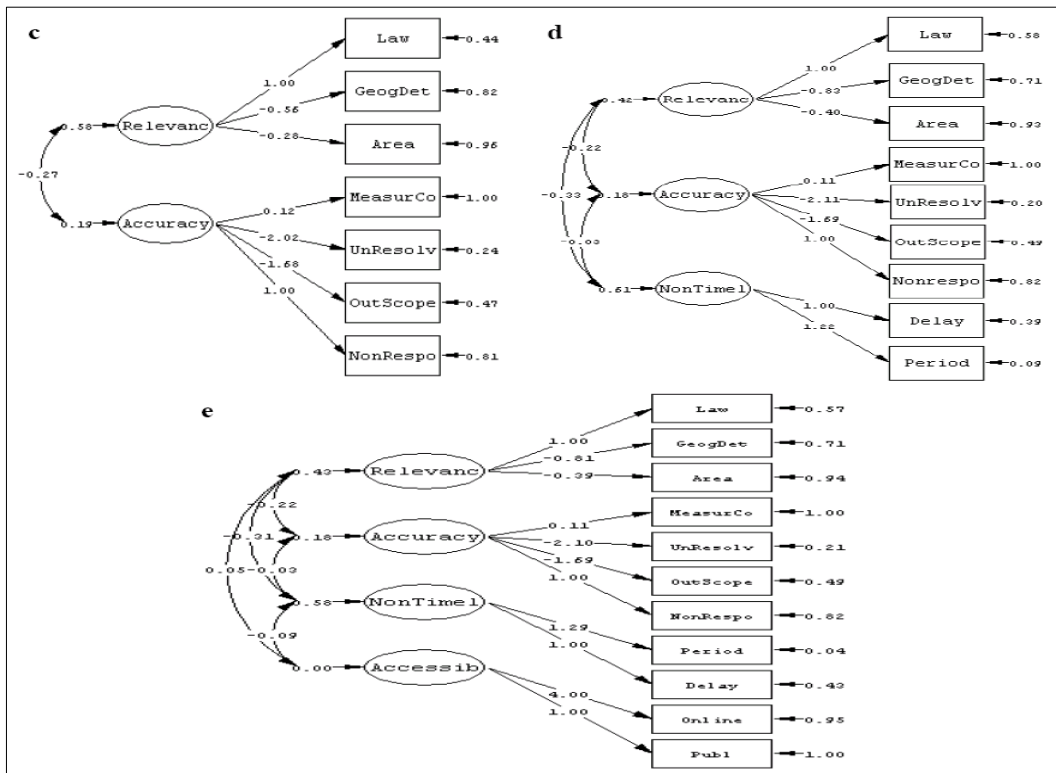
품질 보고서의 1장에서는 일반적인 품질진단의 개요에 대해 서술하고 있다. 통계작성 개요 및 현황과 같은 일반적인 진단 통계의 정보에서부터 품질진단을 어떻게 할 것인가에 대해 서술하고 있다. 진단할 통계의 일반 현황이 주 내용은 아니고 보고서에 따라 차이가 있지만 해당 통계에 대해 잘 모르는 사람이 보고 판단하기에는 통계의 중요도라든지 통계의 현황 및 활용도에 대해서는 자세히 서술되어 있지는 않았다. 또한 품질진단을 하기에 앞서 품질 차원에 대해 정의하는 부분이 있는데 차원이 충분히 진단 목적에 따라 분류되었다고 판단하기는 힘든 부분이 있다. 진단자의 주관적인 부분이 많이 포함되어 되어 있고, 보고서에 따라서는 해당 통계의 차원의 정의보다는 일반적인 차원의 정의에 대해 서술한 보고서도 있다. 품질진단에 있어서 차원의 정의와 분류가 명확해야 진단을 하는데 있어서도 혼동과 누락이 없을 것으로 판단된다.

각 품질 관리 단계별로 점수화에 대한 분석을 해보고자 한다. 품질관리 기반을 측정하는 문제에 있어서 점수화가 가능한 부분은 “제공시기의 정확성 여부”, “인력”, “예산”, “하드웨어”, “소프트웨어”, “이용자 만족도 조사” 등이 있다. 제공시기의 정확성 부분에서는 제공시

Ref.	Notes	Example
B3.5	Unit response rate by sub-groups: business surveys.	(Key Quality Measure)
a	<p>The response rate is a measure of the proportion of sampled units who respond to a survey. This indicates to users how significant the non-response bias is likely to be. Where response is high, non-response bias is likely to be less of a problem than when there are high rates of non-response.</p> <p>The rates opposite can be expressed as a percentage figure by multiplying by 100.</p> <p>NB: There may be instances where non-response bias is high even with very high response rates, if there are large differences between responders and non-responders.</p>	<p>Overall response rate:</p> <p>Unweighted: $\frac{FC + FP + PC + PP}{(FC+FP+PC+PP) + RNU + NR + e(U)}$</p> <p>Weighted: $\frac{\sum_{ieR_0} w_i x_i}{\sum_{ieS} w_i x_i}$</p> <p>Full response rate:</p> <p>Unweighted: $\frac{FC}{(FC+FP+PC+PP) + RNU + NR + e(U)}$</p> <p>Weighted: $\frac{\sum_{ieR_F} w_i x_i}{\sum_{ieS} w_i x_i}$</p> <p>Non-response rate:</p> <p>Unweighted: $\frac{RNU + NR + e(U)}{(FC+FP+PC+PP) + RNU + NR + e(U)}$</p> <p>Weighted: $\frac{\sum_{ieNr} w_i x_i}{\sum_{ieS} w_i x_i}$</p> <p>Where: FC = full period return with complete data; FP = full period return with partial data; PC = part period return with complete data; PP = part period return with partial data; RNU = returned but not used; NR = non-response; U = unknown eligibility; I = ineligible (out of scope); e = estimated proportion of cases of unknown eligibility that are eligible. e can be estimated as: $e = \frac{(FC+FP+PC+PP) + RNU + NR}{(FC+FP+PC+PP) + RNU + NR + I}$</p> <p>$w_i$ = weight for unit i; x_i = value of auxiliary variable for unit i; R_0 = set of all responders; S = set of all eligible sampled units; R_F = set of full responders; Nr = set of non-responders (note that $Nr = S - R_0$).</p>

Accuracy

<그림 1 정확성 관련 무응답 품질 인자>



<그림 2> SEM 분석 결과 도식

기의 기간이 정해져 있으면 100, 제공시기의 오차 부분만큼의 점수화를 할 수 있을 것이다. 인력 부분은 전문 인력의 수로 점수화 할 수 있는데 통계 전문가가 1명 이상이면 100, 통계 전문가가 아니지만 100% 전담 인력이 있으면 90 이런 방식으로 점수화가 가능할 것이다. 또한 통계 교육 이수 정도를 점수화할 수 있는데 3년 이내 전문 통계 교육을 받았으면 100, 5년 이내 통계 교육을 이수 했으면 80 이렇게 점수화 할 수 있다. 예산 부분은 전년대비 예산 증감률을 가지고 점수화를 한다. 작년대비 5% 이상의 예산 증가가 있다면 100, 3% 미만 50, 예산 감소 20 이렇게 점수화를 한다. 하드웨어는 통계 처리 시스템의 보유 여부로 판단한다. 통계 처리 시스템을 자체 보유하면서 자체 운영하면 100, 자체보유하면서 외부 운영이면 80, 외부 보유하면서 자체 운영은 60 등으로 점수화한다. 소프트웨어 사용 부분은 전문 통계 처리 소프트웨어 사용으로 판단한다. 자료처리에 있어서 통계 처리 소프트웨어의 사용은 중요하다. 사무용 소프트웨어인 엑셀로도 어느 정도의 통계 처리는 가능하나 그 정확성에 대해서는 장담할 수 없다. 따라서 전문 소프트웨어를 하나라도 사용한다면 100, 엑셀에서 고급 통계 기능을 사용한다면 80 등으로 산정한다. 이용자 만족도 부분에 대해서는 신뢰도를 표시해야 하고 전체 표본 대상에 대한 표시가 필요하다. 또한 이러한 전체 표본 수 대비 이용자 만족도 조사에 참여한 사람의 수에 대한 점수화가 필요하다.

통계이용실태 부분에서는 “이용 빈도”, “통계자료 이용 목적”, “재사용 의사”가 점수화가 가능할 것으로 보인다. 통계 이용자의 통계 이용 빈도는 연간 사용횟수와 인원에 대한 비율로 점수화가 가능하다. 연간 사용횟수와 인원을 곱해서 연간 총 이용자 횟수와 나누면 비율이 계산된다. 이것을 점수로 사용한다. 통계 이용 목적은 해당 통계의 이용 항목이 있는데 이 항목의 중요도를 설정하여 중요도와 이용 인원을 비율로 계산하여 점수로 사용한다. 재사용 의사도 마찬가지로 이용할 의사 항목의 중요도를 설정하고 이용 인원을 곱하여 비율로 계산하여 점수화하여 사용할 수 있다.

품질 진단에 있어서 수집 자료의 정확성은 가장 중요한 부분이라고 할 수 있다. 이 부분에서는 주로 입력 시스템과 입력된 자료의 정확성 판단에 대한 점수화가 필요하다. 대부분의 진단 보고서를 보면 공통적으로 입력 시스템이 전산화되어 있는 부분은 믿고 넘어가는 경우가 있다. 하지만 입력 시스템이 공통적인 입력 오류를 가지고 있다면 입력하는 데이터 모두를 신뢰할 수 없을

정도로 치명적이다. 따라서 입력 시스템을 맹신하기 보다는 더욱 철저한 진단이 필요할 것으로 판단된다. 소프트웨어의 진단은 직접 입력해보는 것이 가장 좋다. 통계 작성 기관에서 조사된 데이터를 직접 받아서 최소 10번 정도는 직접 입력해보고 오류는 없는지 오류 가능성은 없는지 꼼꼼히 진단해 보아야 할 것이다. 이러한 작업이 어렵다면 전산 전문가의 도움을 받아서라도 진행해야 할 사항은 분명하다. 또 다른 진단 사항은 입력 전산 시스템이 자체 보유, 운영 중이라면 보안에 취약할 가능성이 있다. 만일에 하나 보안상의 문제 발생으로 인하여 데이터에 문제가 생길 수 있다. 이러한 부분은 보안 전문가나 보안 전문 업체에 의뢰하여 분석하고 결과대로 수정할 필요가 있다.

입력 시스템 진단 다음에 중요한 것이 입력된 값을 확인하는 작업일 것이다. 입력 시스템에서 잘 입력된 자료라 하더라도 입력 당시에 확인되지 못하거나 잘못된 값인데 그냥 입력된 값일 수 있다. 보통 조사된 값은 모두 같은 값이 아니고 일정한 변동을 가진 값이기 때문에 잘못 입력된 값을 눈으로 확인할 수는 없다. 또한 입력된 값의 개수가 많은 경우 또한 눈으로 일일이 확인하기란 어려운 일이다. 이러한 경우 자료의 평균이나 분산, 사분위수, 이상점 등의 통계 보고서를 작성하여 확인하는 것이 한 방법이라고 할 수 있을 것이다. 그것이 정말 잘못된 값이 아니더라도 잘못된 값의 후보를 찾아내어 실제 값과 비교하는 방법은 사용할 수 있을 것이다. 식스시그마 소프트웨어나 일반 통계 패키지의 품질관리 부분의 분석 기능을 사용하면 어느 정도 식별 가능한 이상한 값의 후보군을 찾아 낼 수 있다. 다만 이러한 분석 기능은 우리가 품질 진단을 하는데 있어 시간이 많이 걸리거나 찾는 기능이 잘 완비된 소프트웨어를 찾기는 어려울 수도 있다. 따라서 품질진단에서 입력 값의 진단을 수행할 수 있는 별도의 품질 진단 소프트웨어를 제작할 것을 제안한다. 정확성을 진단해야 하는 자료를 입력하면 기초적인 통계 값 출력 이외에 품질에 영향을 미칠 수 있는 값의 후보군을 찾아서 경고를 해줄 수 있는 소프트웨어가 필요하다고 판단된다. 이렇게 하면 데이터 품질에 가장 큰 영향을 미치는 정확성을 어렵지 않게 수행할 수 있을 것이라고 판단된다. 장기적인 품질에 영향을 미치는 요인으로는 응답부담이 있을 수 있다. 응답자는 응답 대내외 환경, 응답문항수, 방문횟수 등에 응답부담을 느끼게 된다. 응답 부담은 응답 품질에 안 좋은 영향을 미친다. 피로도가 높아지고 응답 자체에는 신경을 쓰지 않게 된다. 따라서

이러한 응답 부담을 줄여주는 것이 높은 품질을 얻기 위한 노력이라고 할 수 있다. 이러한 요인에 대해서는 먼저 조사 문항의 개수로 점수화가 가능할 것이다. 응답 문항이 필요에 의해 어쩔 수 없이 많을 수는 있으나 비정상적으로 많거나 할 경우 이를 줄이는 노력을 통해 품질에 힘써야 할 것이다. 두 번째로는 평균 방문 횟수가 있을 것이다. 평균 방문횟수가 비정상적으로 높을 경우 응답자가 의도적으로 응답을 회피하는 건지 아니면 조사 방법에 문제가 있는지 확인할 필요가 있다. 이러한 방문횟수를 줄이는 것 또한 품질을 높이기 위한 기초 노력이 될 것이다.

수집된 자료를 사용자에게 제공하는 것은 사용자들이 자료의 정확성과는 별개로 통계의 품질을 인식하게 되는 중요한 수단이다. 과거에는 인쇄된 매체로 제공하던 것을 최근에는 정보통신의 발달로 인터넷이나 이메일 CD나 DVD, USB 매체로 제공하는 것이 일반화 되었다. 매체의 중요도에 따른 순위를 매기기는 어렵지만 매체 제공의 종류의 개수로 점수화가 가능할 것이다. 종류의 개수가 많아지면 이용자들이 그만큼 편하게 자료를 활용할 수 있는 척도가 될 수 있기 때문이다. 예를 들어 4가지 이상의 종류를 제공하면 100, 3가지 이상은 80 등과 같이 점수화가 가능할 것이다. 제공 매체에 담긴 자료의 내용도 중요할 것이다. 일반적으로 조사된 데이터를 모두 공개하지는 않는다. 데이터에 민감한 자료도 있을 수 있고 개인정보와 사업체 정보 등이 수록될 수도 있기 때문이다. 하지만 이러한 한계점 속에서도 조사된 자료는 사용자에게 가능하면 최대한 제공하는 것이 통계 조사의 목적이 될 것이다. 자료가 불필요하게 제한적으로 제공되면 사용자는 반대로 통계의 품질이 좋지 않다고 생각할 수 있기 때문이다. 따라서 제공 자료의 범위에 대한 점수화가 필요하다. 100% 모든 데이터를 제공하면 100등과 같이 자료 제공 %에 따른 점수화가 가능하다. 영국 통계청의 사례에서처럼 이러한 통계 자료 제공에는 비밀보호 수준이 동반되어야 한다. 또한 통계가 조사된 시점부터 공표된 시점에 대한 시간 간격을 측정하여 얼마나 사용자에게 빨리 그 통계를 사용할 수 있게 해주었는가에 대한 점수화도 필요하다.

<표 5>는 각 단계별 점수화에 대한 항목이다. 보고서를 참조하여 최대한 측정 가능한 지표를 작성하였다. 일부 항목들은 공통되지 않고 상세한 정보가 없어 지금의 보고서 단계에서는 측정이 불가능하였다. 앞의 점수화 방안에 의해 “농림업생산지수”의 품질 진단을 점수화하고 비교하였다. 전반적으로 낮은 점수를 받았으나

현재 상태에서는 이 점수는 큰 의미를 갖지는 않는다. 정보가 없는 항목은 제외를 했지만 점수화를 위해 선정했던 항목에 대한 자세한 정보가 없는 상태에서 점수화를 하기엔 힘든 부분이 있었다. 이렇게 점수화를 하면서 알게 된 것은 점수가 큰 의미는 없지만 전반적으로 개별 품질 진단에 대한 자체의 품질에 대한 어느 정도의 평가가 가능하다는 것이다. 어느 품질 진단 보고서에는 있는 내용인데 다른 보고서에는 없는 항목도 있다. 세부적으로 들어가 보면 측정해야 하는 항목임에도 불구하고 보고서에는 자세히 서술되어 있지 않다거나 누락된 경우가 있었다. 즉 점수가 낮은 이유가 품질 진단의 자체 점수가 낮아서 일수도 있고 자세한 정보를 기록하지 않은 보고서의 품질을 대변할 수도 있다는 것이다. 향후에는 이러한 품질 진단 보고서를 정형화하여 누락되거나 상세 정보가 없는 경우를 제거하여 품질 진단 자체에 대한 평가만 가능하게 하여야 할 것이라고 생각된다.

이러한 평가를 제대로 하게 하려면 일단 평가를 하려는 항목에 대해 필수적으로 보고서에 기재하여야 하고 상세한 항목을 연구하여 보고서 작성 항목에 넣어야 할 것이다. 대부분의 보고서의 상세 내용이 온라인이나 제출된 문서의 정보를 그대로 가져오거나 절대적으로 의존하는 경우가 많은데 이렇게 진단을 하게 되면 상세한 정보를 듣지 못해 잘못된 진단을 내릴 수 있는 가능성을 가지게 된다. 상세한 정보, 정확한 정보를 알 필요가 있을 경우엔 담당자나 기타 방법을 통해 알아내서 정확하게 기록하게 할 필요가 있다고 판단된다. 아래는 각 품질 진단 별 보고서의 목차이다. 보고서의 목차를 보면 정형화의 필요성과 어떠한 품질의 목적에 집중하여야 할지에 대한 방향성을 볼 수 있다고 판단된다.

5. 결 론

본 연구에서는 통계의 광의의 품질이란 무엇인가에 대해 알아보고 품질 진단이란 무엇인가에 대해 정의하였다. 또한 통계청이 품질진단을 하는데 사용하는 품질의 정의에 대해 알아보고 각 품질 진단의 각 단계에 대해 알아보았다. 품질진단이 어디서 출발하였는지 역사에 대해 알아보고 각국이 품질진단을 어떻게 시작하여 실시하고 있는지 알아보았다. 그리고 해외에서는 품질진단을 개선하기 위한 연구로 어떠한 것을 실시하여 개선에 반영하였는지 알아보고 국내의 연구 성과에 대해 알아보았다. 최근의 통계청 품질진단의 방향과 방법, 절

<표 5> 품질 지표 점수화 방안

진단 영역	항목	내용	예시
품질 관리 기반	제공시기의 정확성 여부	제공 시기 확정 여부에 따른 점수화	월 일 확정 게시 시 100
	통계 전문 인력 여부	통계 전문 인력이 전담하고 있는가에 대한 점수화	통계 전문 인력이 업무 전담 : 100 비전문 인력 업무 전담 100% : 100
	통계교육 이수 여부	통계 전문 교육 이수 여부에 따른 점수화	1년 이내 전문 교육 이수 : 100 2년 이내 전문 교육 이수 : 80
	통계 예산	통계 관련 예산 증액 여부	전년 대비 5% 이상 증가 : 100 전년 대비 3% 미만 증가 : 80
	하드웨어	통계 처리(입력, 분석) 시스템 보유 여부	자체 보유, 자체 운용 : 100 자체 보유, 외부 운용 : 80
	소프트웨어	통계 처리 시 전문 통계 소프트웨어 사용 여부	전문 소프트웨어 사용 : 100 엑셀 통계 기능 사용 : 80
통계이용실태	통계 이용 빈도	연간 이용 빈도에 따른 점수화	연간 사용 횟수 x 이용 인원 비율
	통계 자료 이용 목적	이용 목적의 중요도에 따른 점수화	이용 목적의 중요도 가중치 x 이용 인원
	이용자 만족도	통계 이용자의 만족도 점수	종합만족도(이미 존재)
	제사용 의사	제사용 의사의 중요도에 따른 점수화	제사용 의사 중요도 가중치 x 응답 인원
수집자료의 정확성	입력 시스템 점검	입력 시스템 점검 여부에 따른 점수화	보안+ 정기 점검 + 업데이트 : 100 보안 + 정기 점검 : 80
	입력된 데이터	품질 분석 결과에 따른 점수화	별도의 프로그램을 이용하여 점수화
	입력 데이터 내검	데이터 내검 방법에 대한 점수화	5단계 이상 내검 : 100 4단계 이상 내검 : 80
	내검 수준	내검 정도에 따른 점수화	수정된 자료의 수 / 조사된 데이터의 총수
	무응답 대체	무응답 대체 정도에 따른 점수화	대체된 데이터의 총 수 / 모든 최종 데이터의 수
	수집된 자료 제공	제공 매체 종류	제공 매체 종류에 의한 점수화
자료 공표	자료 공표 범위	자료 공표 범위에 따른 점수화	자료 공표 %에 따른 점수화 셀 단위 카운트
	비밀보호 수준	비밀보호 기법 적용 여부에 대한 점수화	감춰진 셀의 비율
	공표시기	통계 조사가 이루어진 시기와 공표시기와 시간차	통계마다의 차이가 있지만 적절한 시기를 고려해서 점수화

차 등에 대해 알아보았으며 2008년에 실시된 품질진단의 몇 가지 실제 사례를 살펴보았다.

국가 통계가 아닌 일반 제조업 등의 산업에서의 품질이란 제품의 성능이나 제품의 결합 등의 수치 즉 통계를 가지고 측정하며 품질 진단을 통한 품질 향상 노력을 한다. 그러나 통계에 있어서의 품질이란 정의하기도 어렵고 측정하기도 어렵다는 것을 해외 연구 사례나 국내 사례에서 쉽게 알 수 있다. 통계청에서 실시하는 품

질진단도 가장 중요하다고 생각되는 6개의 차원에 대해서 할 뿐이지 정말 제대로 세밀하게 품질진단을 한다면 그보다는 많은 차원에 대해 품질을 정의하고 각각의 차원에 대한 항목별로 품질진단을 하여야 할 것이다. 그러나 한정된 시간에 한정된 자원을 가지고 많은 수의 통계를 심층적으로 하는 것은 현실적으로 불가능 할 것이다.

품질진단이 어렵고 힘든 작업이라고 해서 통계청이

<표 6> 품질 지표 점수화

진단 영역	항목	내용	점수
품질 관리 기반	제공시기의 정확성 여부	매년 8월경 제공	80
	통계 전문 인력 여부	전문 인력 미비에 부분 참여	60
	통계교육 이수 여부	3년 이내 8회 40일 이수	60
	통계 예산	별도 예산 X	0
	하드웨어	농업사업통합정보시스템 Agrix	100
	소프트웨어	농업사업통합정보시스템 Agrix	100
	이용자 만족도 조사	66/257(제공 받은 통계 사용자 리스트)	25
	만족도 조사 조사 방법	전화/e-mail	40
통계이용실태	통계 이용 빈도	연 2회 이상 비율	60.61
	통계 자료 이용 목적	정책진단/연구 학습	84.85
	이용자 만족도	3.51(보고서 참조) / 5	70.2
	재사용 의사	그렇다 응답 이상	86.36
수집자료의 정확성	입력 시스템 점검	자료 없음	X
	입력된 데이터	자료 없음	X
	입력 데이터 내검	자료 없음	X
	내검 수준	자료 없음	X
	무응답 대체	자료 없음	X
	조사문항수	자료 없음	X
	평균 방문 횟수	자료 없음	X
수집된 자료 제공	제공 매체 종류	PDF, WEB	60
자료 공표	자료 공표 범위	자료 없음	X
	비밀보호 수준	자료 없음	X
	공표시기 차이	자료 없음	X
총점	827.02		
평균	63.61		

나 통계생산 기관이 통계의 품질 향상에 소홀할 수는 없을 것이다. 본 연구에서 찾아낸 어려움은 첫째, 각각의 통계에 대해 품질진단을 개별적으로 하다 보니 진단 마다의 차이가 많이 발생하고 있음을 알 수 있다. 결론적으로 보면 시간과 인력을 많이 들여서 하는 품질 진단에 대한 품질을 보장 못하는 결과가 얻어지게 되는 것이다. 품질 진단을 하는 개별 개체가 생각이 다르고 방향이 달라서 차이가 발생하는 것 같다. 다시 말하면 품질 진단에 개별 진단 주체의 주관적인 생각이 너무 많이 반영되는 것과 같은 문제점이 발생하는 것이 보인다.

두 번째로는 품질에 있어서 수집 자료의 정확성이 가장 중요한 차원이라고 많은 기존 연구에서 밝혔음에도 불구하고 정확성에 대한 측정을 대부분 입력 시스템의 전산화나 눈에 띄는 이상치가 없으면 그냥 넘어가는 경우가 종종 있다는 것이다. 입력 시스템이 전산화되어 있는 것과 수집 자료의 정확성과는 크게 관련이 없다. 입력이 정확하다는 한 가지 장점이 있는 것이지 조사된 자료 자체가 정확하다는 것과는 별개의 것일 것이다. 따라서 해외 연구 자료에 의하면 정확성을 측정하기 위해 더 많은 측정 항목을 개발하고 있는 실정이다. 이러

한 자료의 정확성은 사실 진단 주체들이 일일이 분석하고 자료의 정확성을 판단하기에는 너무 많은 시간과 인적 자원 및 전문 인력이 필요하다. 따라서 진단 주체에 이러한 몫을 남길 것이 아니라 품질진단을 위한 별도의 자료의 정확성 판단 품질 진단 소프트웨어를 장기적으로 설계하여 개발할 필요가 있다. 이러한 소프트웨어는 기초통계가 외에 사업체나 가구냐에 따른 자료의 성격에 따라 경제 자료인 경우 간단한 시계열 분석 등을 포함해야 한다. 이러한 소프트웨어를 사용하여 진단 주체들은 많은 시간을 할애하지 않아도 수집 자료를 간단하게 DB 형태나 자료 형태 그대로 입력하여 품질에 대한 분석을 가능하게 하여야 한다.

세 번째는 품질 진단이 통계 품질 자체보다는 정책사용 가능성이나 이용자 측면에 대한 부분이 더 많다는 것이다. 입력 자료의 정확성 부분이 다른 부분보다는 더 중요하다는 것은 아무도 이의를 제기하기는 어려울 것이다. 자료 자체의 품질이 보장되지 않으면 정책에 잘못 사용될 가능성, 이용자들이 잘못된 사용으로 인해 혼란을 야기할 가능성이 너무나도 많다. 그러나 품질 진단 매뉴얼에는 수집 자료의 정확성을 세밀하게 분석하고 확인할 수 있는 설계가 부족한 것 같다. 해외 사례에서 보았듯이 자료의 기초 통계는 물론 자료의 수정, 응답자의 비율, 자료의 대체 등 품질에 영향을 미칠 수 있는 많은 자료의 수집부터 가공의 단계를 측정하고 수

치화 할 필요가 있다. 본 연구에서 제시한 몇 가지 수치는 공통적인 것이지만 통계의 에 따라 세밀한 측정 도구를 설계하여 제시해야 할 것이다.

참고문헌

- [1] 품질관리과(2008), 「“기업결합동향” 2008년 정기 통계품질진단 연구용역 최종결과보고서」, 통계청.
- [2] 품질관리과(2008), 「“농림업생산지수” 2008년 정기 통계품질진단 연구용역 최종결과보고서」, 통계청.
- [3] 품질관리과(2008), 「“문화재 관리현황” 2008년 정기 통계품질진단 연구용역 최종결과보고서」, 통계청.
- [4] 품질관리과(2008), 「“재래시장 실태조사” 2008년 정기 통계품질진단 연구용역 최종결과보고서」, 통계청.
- [5] 품질관리팀(2004), “외국의 품질관리”, 통계청.
- [6] Giovanna Brancato, Giorgia Simeoni(2008), “Modelling Survey Quality by Structural Equation Models”, Istat.
- [7] EUROSTAT(2005), “Standard quality indicators”, Quality in Statistics.
- [8] Karen Dunnell(2007), “Guidelines for Measuring statistical Quality”, Office for National Statistics.

2010년 2월 22일 접수, 2010년 6월 8일 1차 수정, 2010년 6월 14일 2차 수정, 2010년 6월 17일 채택