

이러닝수요부문 사업체실태조사를 위한 표본설계

김혜중¹ · 곽화륜²

¹동국대학교 통계학과 · ²동국대학교 통계정보기술연구소

접수 2013년 5월 10일, 수정 2013년 6월 12일, 게재확정 2013년 6월 20일

요약

이러닝 (e-learning) 산업통계는 이러닝산업 전반에 걸친 수요와 공급을 망라한 실태조사통계로 2004년 이후 정보통신산업진흥원에 의해 매년 발표되고 있다. 한국표준산업분류가 2008년 개정 (제 9차)됨에 따라, 이러닝 수요부문의 사업체조사에 사용해진 표본설계 (종사자규모·산업분류별 층화추출)에 대한 개선이 필요하게 되었다. 본 논문은 사업체조사의 목표모집단을 개정된 한국표준산업분류에 따라 종사자규모·산업분류별로 층화하고, 각 층에 부여된 목표변동계수값을 만족시키는 멱배분법의 승수를 모의실험으로 찾아서 층화별로 표본의 크기를 결정하였다. 이와 더불어 본 연구에서 고려한 표본가중치 계산, 그리고 가중치를 반영한 모수추정량과 추정오차는 기존의 조사에서 사용된 기술통계적 분석을 벗어나 업종별과 종사자 규모별 추정과 추정의 정도에 대한 평가를 가능하게 하였다.

주요용어: 멱배분, 이러닝산업통계, 층화, 표본설계, 한국표준산업분류.

1. 머리말

이러닝산업실태조사는 통계청 승인통계로서 산업통상자원부 (구 지식경제부)에서 정보통신산업진흥원에 위탁하여 생산하는 통계이다. 2003년 이러닝산업통계에 대한 조사체계를 수립하고, 2004년에 출범한 이러닝산업실태조사 (Ministry of Knowledge Economy·National IT Industry Promotion Agency, 2012)는 국내 이러닝 산업의 전반적인 규모와 현황을 파악하는 정책의 기본 자료를 제공한다. 또한 이 조사는 미래 신성장동력 산업에 대한 산학연구 및 기업들의 신산업 진입 등에 기초자료를 제공할 뿐 아니라, 국내 이러닝산업 시장을 전망하여 이러닝 산업 경쟁력을 강화하고 건전한 수요창출을 통한 이러닝의 확산에 도움을 주는 통계를 생산하고 있다. 실태조사에서는 국내 이러닝산업 전반에 걸친 공급과 수요부문을 대상으로 하고 있으며, 공급부문은 이러닝산업에 종사하는 사업체를 네 가지 사업 분야인 콘텐츠, 솔루션, 인프라, 교육서비스사업별로 층화하여 조사하고 있으며, 수요자 부분은 네 가지 수요그룹 (개인, 사업체, 정규교육기관, 정부 및 공공기관)으로 나누어 각각 조사하고 있다.

본 통계가 우리나라의 이러닝산업을 제대로 반영하고 있는지에 대한 포괄적인 점검결과, 이러닝 수요자부문의 사업체조사에서 사용된 산업별 층화는 제 9차 한국표준산업분류 (KSIC9; Korea National Statistics Office, 2008)와 일치하지 않고, 추정량의 정확도에서도 만족스럽지 못한 것으로 나타났다 (Ministry of Knowledge Economy, 2012). 즉, 현행조사에 사용된 표본설계는 KSIC9 (9th Korean Standard Industrial Classification)에서 상이한 업종으로 분류한 광업과 제조업을 동일한 업종층 (stratum)으로 통합한 표본 설계이다. 그 결과 현행조사는 산업별 중요정보에 대한 분석 및 비교에 필요한 통계정보를 충분히 제공하지 못할 뿐 아니라, 모수추정의 정확도에서도 통계청 승인통계의 권고기

¹ 교신저자: (100-715) 서울 중구 필동, 동국대학교 통계학과, 교수. E-mail: kim3hj@dongguk.edu

² (100-715) 서울 중구 필동, 동국대학교 통계정보기술연구소, 연구원.

준 (전국기준 5% 이내의 변동계수 (CV))를 만족시키지 못하는 것으로 나타났다. 또한 기존의 표본설계는 층별 표본배분에 산업별/사업체별로 상이한 이러닝 수용비율을 고려하지 않아, 결국 비현실적인 가중치를 사용하여 모수를 추정하고 있는 것으로 나타났다.

본 논문은 이러닝 수요부문사업체조사 (이하의 문장에서 이러닝사업체조사로 사용)통계의 품질향상을 위해 기존의 표본설계와는 달리 KSIC9의 대분류체계 (18개 업종)에 따라 광업과 제조업을 분리하여 목표모집단을 층화하고, 층별 표본배분에는 산업 및 사업체별 이러닝 수용비율을 반영시킨 새로운 표본설계를 제안하고자 한다. 또한 본 연구에서 제안한 표본설계가 국가승인통계의 기준에 걸맞은 정확도를 가진 이러닝사업체조사통계의 생산에 적합한 것임을 모의실험을 통해 보이고, 특이치 및 무응답처리, 통계의 시계열 유지 등 기타 이러닝사업체조사통계 생산에서 유의할 사항들을 제안하고자 한다.

2. 현행 수요자부문의 사업체조사에 대한 분석

2.1. 현행표본설계개요

현행 이러닝산업실태조사는 표본설계 측면에서 공급자 부문과 수요자 부문으로 구분되며, 수요자부문의 표본추출틀은 네 종류로 구분된다. 먼저 개인부문은 ‘지역별 전화번호부’ 리스트, 사업체부문은 2010년 통계청 ‘사업체기초조사’ 명부, 정규교육기관부문은 ‘한국교육개발원 2010 정규교육기관’ 명부, 정부 및 공공기관부문은 ‘정부기관 및 정부산하 공공기관’ 명부를 각 조사의 표본추출틀로 사용하고 있다 (Ministry of Knowledge Economy, 2012).

특히 본 연구의 대상인 이러닝사업체조사에 사용되고 있는 표본설계는 상시근로자 1인 이상인 전 업종 사업체를 대상으로 하고 있으며, 국제기관 및 가사서비스업은 조사에서 제외하고 있다. 표본의 크기는 7,164개로 정하여, 추정 (사업체의 이러닝 총 지출액 추정)의 정도 95% 신뢰도에서 허용오차는 $\pm 5\%p$ (전년도 지출총액 대비기준)이내가 되도록 하고 있다. 이를 위해 현행의 표본설계는 산업 및 규모별 층화표본추출법을 사용하고 있으며, 층별 표본배분은 승수가 $p=.5$ 인 멱배정 (power allocation)에 의해 표본을 산업 및 규모층 별로 할당하여 300인 미만 사업체에 지나치게 많은 사업체가 할당되는 문제를 해결하고 있다 (Ministry of Knowledge Economy·National IT Industry Promotion Agency, 2012).

2.2. 통계품질의 문제점 및 개선사항

‘이러닝산업실태조사’가 개시된 2003년도의 이러닝산업과 현재의 이러닝산업은 그 범위와 수요자 및 공급자의 분포가 매우 다른 양상을 보일뿐 아니라, 2008년도에 들어서 한국표준산업분류체계가 KSIC8에서 KSIC9로 개정된 실정이다. 따라서 이러닝사업체조사통계의 품질제고를 위해서는 조사모집단에 대한 재설계가 기본적으로 필요한 상황이다. 또한 이러닝사업체조사통계에 대해 2010년도에 실시된 통계품질진단결과 (Ministry of Knowledge Economy, 2012)와 본 연구에서 실시한 포괄적인 점검결과를 종합하면, 현행의 표본설계에서 다음과 같은 사항의 개선에 대한 방안마련이 필요한 것으로 나타났다. 첫째로 이러닝사업체조사통계는 국가승인통계가 요구하는 통계의 정확도 (전국기준 5% 이내의 변동계수)를 충족하지 못해, 이를 해결하기 위한 방안마련이 필요한 것으로 나타났다. 둘째로 상이한 산업분류 (KSIC9 기준)에 속한 광업과 제조업체들을 동일업종 층으로 통합하여 조사하고 있어, KSIC9의 대분류체계 (18개 업종)에 맞게 광업과 제조업을 분리하여 조사하는 방안마련이 필요한 것으로 나타났다. 셋째로, 표본배분 방법에 사업체의 이러닝 수용비율을 반영시킬 수 있는 방안마련이 필요한 것으로 나타났다. 따라서 본 논문은 이와 같은 문제들을 해결할 수 있는 새로운 표본설계의 제안과 표본조사수행시 고려하여야 하는 특이치 검출 및 처리, 무응답처리, 통계의 시계열 유지 등 제반 문제점들에 대한 인식과 그 해결방안의 마련에 관한 내용을 순차적으로 다루며, 이러한 내용들은 이러닝사업체

조사통계의 품질향상에 기여할 것이다.

3. 표본설계

3.1. 표본설계의 특징

이러닝산업실태조사를 위해 본 연구에서 새로이 제안한 표본설계는 다음과 같은 특징을 가진다. 첫째, 한정된 조사비용과 통계의 시계열성 유지를 고려하여, 전체표본 수는 기존 조사의 표본 수와 같이 7,200개 내외의 표본을 유지하였다. 둘째, 본 조사의 표본추출법은 층화추출법이며, 층화변수로는 과년도와 같이 ‘업종’과 ‘종사자 규모’로 하였다. 셋째, 조사모집단은 통계청 2010년 전국사업체 기초조사명부에 있는 상시근로자 1인 이상 사업체로 정의한다. 또한 조사대상 범위는 KSIC9의 대분류 기준 18개 업종을 원칙으로 하고, 상대적으로 사업체수가 많은 제조업의 경우는 중분류(식료품 제조업 등 24개 업종)까지를 고려하였다. 단, 산업 활동과 관련이 없는 3개 업종인 ‘공공행정, 국방 및 사회보장 행정’, ‘가구 내 고용활동 및 달리 분류되지 않은 자가소비 생산활동’ 및 ‘국제 및 외국기관’은 조사에서 제외시켰다. 넷째, 본 조사의 표본설계에 사용되는 층화변수인 종사자(상시근로자)규모는 연차비교를 위하여 과년도와 동일 기준인 1~9인, 10~49인, 50~299인, 300인 이상과 같이 4개의 범주로 분류하고, 이를 토대로 모집단을 업종 및 종사자규모별로 층화하였다. 다섯째, 2011년도 조사자료에서 이러닝 지출액과의 상관계수가 .664 (p -값<.0001)로 높게 나타난 종사자 규모를 층화변수로 사용하여, 사업체의 이러닝 지출액과 직접 관련이 있는 수용비율을 표본설계에 자연스럽게 반영시켰다(Figure 3.1 참조). 여섯째, 표본배분법의 가중치 계산에 이러닝사업체조사의 중요변수(key variable)인 이러닝 지출액의 과년도 변동계수 또는 분산을 사용하는 것이 최선이다. 그러나 전년도 표본설계의 업종층화(17개 업종으로 층화)를 통해 조사된 층별 이러닝 지출액자료를 새로운 표본설계의 업종층화(41개 업종분류)에 맞게 자료를 변환시키는 것은 현실적으로 불가능하다. 따라서 새로운 표본설계에서는 전국사업체 기초조사통계로부터 정확한 자료의 수집이 가능한 동시에 중요변수와 상관이 높은 종업원 수에 대한 층별 변동계수를 계산하여 표본배분방법의 가중치로 사용하였다. 마지막으로, 각 층(업종 및 종업원규모층)에 할당될 표본의 크기는 맥배분법을 사용하였으며, 여기서 사용되는 승수값은 조사의 정확도기준에 의해 선택하였다.

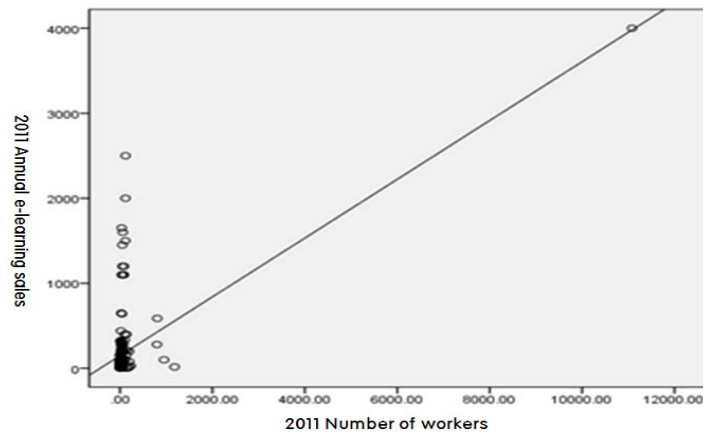


Figure 3.1 Plot of annual e-learning sales versus number of workers
(source: 2011 e-learning industry status survey)

3.2. 모집단 분석

2010년 통계청 사업체기초조사는 상시근로자 1인 이상 사업체에 대해 전수조사를 실시하였다. 이 조사로부터 파악된 전체 사업체들 중에서 산업 활동과 관련이 없는 3개 업종 (공공행정, 국방 및 사회보장행정, 가구 내 고용활동 및 달리 분류되지 않은 자가소비 생산활동, 국제 및 외국기관)에 속한 것을 제외한 나머지는 모두 3,343,541개 사업체로 파악되었다. 이런 방법으로 파악된 사업체집단이 새로운 표본설계의 조사모집단이 되었다 (Table 3.2 참조). Table 3.2에 의하면 조사모집단을 KSIC9의 대분류 기준인 18개 업종별로 분류하였을 때, 제조업, 도매 및 소매업, 운수업, 그리고 숙박 및 음식점업중이 타 업종들에 비해 상대적으로 규모가 큰 부모집단을 이루는 특징을 보이고 있다. 또한 이들 업종을 종업원 수의 분포로 비교하면, '9인 이하'의 사업체가 제조업 (80.9%), 도매 및 소매업 (96.4%), 운수업 (96.8%), 숙박 및 음식점업 (97.9%)과 같은 비율을 차지하고 있다. 따라서 조사모집단에 대한 표본설계는 먼저 '9인 이하'사업체의 비중이 상대적으로 낮은 제조업을 대상으로 중분류까지 층을 나누어 규모가 큰 부모집단 문제를 부분적으로 해결하고, 나머지 업종들이 보이는 종업원규모의 쏠림현상은 멱배분법으로 해결하는 방법의 사용이 필요한 것으로 나타났다.

3.3. 층화

본 표본설계는 층화표본방법을 사용하며, 조사모집단의 층화에는 변수인 KSIC9체계상의 업종과 종사자규모를 층화변수로 선택하였다. 조사모집단 층화에서 고려된 사항들은 다음과 같다. 첫째로 2004년~2012년에 걸쳐 축적한 이러닝수요사업체조사통계의 시계열유지를 가능하게 하고, 둘째로 각 업종별·종사자규모별 통계작성 뿐 아니라 세부항목인 업종별 및 종사자규모별 추정 및 비교를 할 수 있도록 고려하였다. 업종별 층화는 KSIC9의 대분류 기준 18개 업종을 원칙으로 하고, 상대적으로 부모집단의 크기가 큰 제조업의 경우에는 중분류 (24개 업종)까지 세분화하여 층화하였다. 각 업종별로는 통계청 2010 '사업체기초조사' 명부에 명시된 종업원 수를 크기측도로 하여, 종사자 규모가 9인 이하, 10인~49명, 49인~299명, 300명 이상인 계급 (4개)으로 나누어 시계열유지 및 공급자 부문조사와의 비교가 가능하도록 층화하였다. 따라서 조사모집단은 총 408개 (17×24)의 층으로 구성되며, 타 분야에서 실시되는 사업체 (또는 기업체조사)조사의 조사모집단 층화와는 차이가 있다 (Kim 등, 2010; Heo와 Chang, 2010).

3.4. 표본크기 결정 및 배분

전체 표본수는 과년도와 표본설계와 동일하게 7,200개 내외의 표본을 유지하도록 하여, 사업체조사 통계의 시계열성 유지 및 비교를 가능하게 하였다. 과년도 표본설계와 같이 멱배정법으로 총 표본사업체 (7,200개 내외)를 업종별·종사자규모별로 할당하도록 설계하였으며, 멱배정식은 식 (3.1)과 같다. 식 (3.1)에서 사용된 할당가중치의 계산에는 각 층의 이러닝 지출액과 높은 양의 상관 (상관계수=.664, p -값<.0001)을 가진 종사자 수의 변이계수가 사용되었다. 이와 같이 표본표준편차 대신 변이계수를 사용한 식을 적용하는 이유는 층에 따라 종사자 규모의 차이가 매우 큰 부모집단들의 특성 (Table 3.1의 모표준편차 참조)을 할당가중치 계산에 반영하기 위함이다 (Ministry of Knowledge Economy·National IT Industry Promotion Agency, 2012). 참고로 멱배정법의 할당가중치 계산에는 이러닝사업체조사의 중요변수인 이러닝 지출액의 과년도 변동계수 또는 분산을 사용하는 것이 최선이다. 하지만 전년도 표본설계의 업종층화 (17개 업종 층화)가 새로운 표본설계의 업종층화 (41개 업종분류)에 대응되지 않아 층별 이러닝 지출액에 대한 전년도 자료를 새로운 표본설계의 업종층화 (41개 업종분류)에 맞게 변환시켜 할당가중치를 구하는 방법은 본 연구에서 적용될 수 없었다.

추정량의 정도는 각 업종별 모집단 사업체 규모에 따라 표본오차의 절대량으로 목표오차를 설정하기 보다는 상대표본오차 개념인 변동계수를 사용하는 것이 바람직하다고 판단하였다. 이에 종사자수의 변동계수를 사용하여 각 업종별 목표오차를 다음과 같이 설정하였다. 목표오차의 관리를 위해, 전체 산업은 변동계수가 국가승인통계기준인 5% 이내 이어야 하며, 대분류업종 층은 10%내외, 제조업의 중분류업종 층은 25%내외의 변동계수를 가지는 표본설계가 되도록 하였다. 이렇게 정해진 목표오차를 충족시키기 위해 각 층을 독립적인 모집단으로 보고 각 층별로 요구되는 목표정도를 포괄적으로 만족시킬 수 있는 최적의 역배정식 ($p=0.4$ 인 식 (3.1))으로 업종별·종사자규모별 표본수를 할당하였다.

$$n_h = n \times \left(\frac{(N_h S_h)^p}{\sum_{h=1}^L (N_h S_h)^p} \right) \approx n \times \left(\frac{(N_h C_h)^p}{\sum_{h=1}^L (N_h C_h)^p} \right) \quad (3.1)$$

여기서 p = 역배정승수, n = 총 표본규모인 7,200사업체 내외, $n_h = h$ 층 (i 번째 업종 · j 번째 종사자 규모층)의 표본크기, N_h 는 h 층의 부모집단크기, $S_h = h$ 층 종사자 수의 표본표준편차, $C_h = h$ 층 종사자 수의 변동계수이다. 최적의 역배정을 선택하기 위해 식 (3.1)에 나타난 역배정 승수 (power) p 를 0.1씩 증가시키면서 모의실험을 실시하였다. 주어진 p 값을 식 (3.1)에 적용하여 각 층의 표본수를 할당하고, 표본추출틀 (사업체기초조사의 사업체 명부)모의표본추출실험 (100회 반복실험)으로 표집된 기업체의 종업원 수로 부터 계산된 전체산업 및 각 업종별 변동계수값 (모의표본 100회 평균값)을 계산하였다. Table 3.1은 이러한 모의표본실험에서 업종 층별 변동계수값들이 목표오차에 근접한 경우 ($p=0.4, 0.5, 0.6$)을 나타내었다.

Table 3.1에서와 같이 역배정식인 식 (3.1)에 $p=0.4$ 을 사용하여 표본을 할당할 경우, 업종별 변동계수값들이 설계된 목표오차를 대부분 만족시키는 것으로 나타났다. 이와는 달리 다른 승수값을 사용한 역배정식으로 표본할당을 하면, 전국기준 변동계수 ($p=0.5$ 인 경우 1.91%, $p=0.6$ 인 경우 3.28%)는 목표오차 (5% 이내)를 만족시키지만, 각 업종별 변동계수가 정해진 목표오차 (대분류업종 층은 10%내외, 제조업의 중분류업종 층은 25%내외의 변동계수로 정함)는 만족시키지 못하는 것으로 나타났다. Table 3.2는 $p=0.4$ 인 최적의 역배정식을 사용하여 층별로 할당시킨 표본현황을 나타낸다. 최종할당에 있어서, 특정 층의 부모집단 사업체수가 작아 역배정식에 의해 표본할당이 안 된 경우에는 전수를 할당시켰다. 이에 따라 Table 3.2는 실제표본의 할당과정에서 발생된 전수 할당으로 인해 총 표본의 크기는 7,200개 보다 많은 7,204개 사업체로 최종할당 된 표본의 현황을 나타낸다. 이를 종합하면, 업종별·종사자규모별 조사모집단의 규모와 본 연구에서 사용한 표본설계에 의해 최적배분법으로 할당된 업종별·종사자규모별 표본의 수를 나타내고 있다. 또한, 이 값들은 표본계획에서 사용한 목표오차를 대부분 만족시키고 있으며, 추정된 전국기준 변동계수는 1.82%로 계산 되었다. 따라서 본 연구에서 제안된 표본설계는 국가승인통계의 정확도 기준 (5% 이내의 변동계수)을 충분히 만족시키는 정확도를 가진 것임을 나타낸다.

3.5. 표본추출

표본추출틀은 2010년도 기준으로 통계청에서 조사한 ‘사업체기초조사’로 부터 얻은 사업체 명부를 사용하였다. 단, 산업 활동과 관련이 없는 3개 업종인 ‘공공행정, 국방 및 사회보장행정’, ‘가구 내 고용 활동 및 달리 분류되지 않은 자가소비 생산활동’ 및 ‘국제 및 외국기관’은 표본추출틀에서 제외시킨다. 총 3,084,336개 사업체로 이루어진 표본추출틀을 업종별·종사자규모별로 층화하고, 역배정법에 의해 할당된 표본의 크기를 나타낸 Table 3.2의 표본현황대로 각 층에서 표본을 추출한다. 이를 위해 해당 층에 속한 사업체들을 종사자 수의 크기순으로 정렬한 후 계통추출법을 사용하여 표본을 추출하였다. 이때 무응답이 발생할 경우를 대비하여 층별로 표본대체에 사용될 예비표본 (통상 할당된 표본수의 2배)도 선정이 가능하다. 하지만 전수층의 경우 무응답이 발생하면 이를 대체 할 수 없으므로 최대한 무응답을 줄일 수 있도록 해야 하며, 무응답이 발생할 경우는 가중치를 조정해준다 (Lee와 Kang, 2012).

Table 3.1 Coefficient of variations (CV) and standard deviation (sd)

KSIC9 Section and Division Code	power			sd
	$p = 0.4$	$p = 0.5$	$p = 0.6$	
	CV (%)	CV (%)	CV (%)	
A. Agriculture, forestry and fishing	5.32	7.19	8.08	26.21
B. Mining and quarrying	8.55	7.01	5.94	38.46
C. Manufacturing	4.12	4.36	9.50	109.03
10 Manufacture of Food Products	4.71	5.68	8.96	20.59
11 Manufacture of Beverages	6.94	15.90	25.87	39.87
12 Manufacture of Tobacco Products	16.54	NA	NA	153.59
13 Manufacture of Textiles, Except Apparel	4.33	5.70	6.57	17.90
14 Manufacture of wearing apparel, Clothing Accessories and Fur Articles	5.09	4.69	6.14	20.00
15 Tanning and Dressing of Leather, Manufacture of Luggage and Footwear	6.41	9.77	6.82	16.71
16 Manufacture of Wood Products of Wood and Cork ; Except Furniture	6.61	9.15	8.21	12.43
17 Manufacture of Pulp, Paper and Paper Products	4.40	6.74	5.69	28.74
18 Printing and Reproduction of Recorded Media	6.12	6.15	6.81	9.06
19 Manufacture of Coke, hard-coal and lignite fuel briquettes and Refined Petroleum Products	10.43	10.73	9.02	213.00
20 Manufacture of chemicals and chemical products except pharmaceuticals, medicinal chemicals	4.69	7.78	5.34	52.03
21 Manufacture of Pharmaceuticals, Medicinal Chemicals and Botanical Products	8.79	9.42	8.37	59.16
22 Manufacture of Rubber and Plastic Products	5.26	5.42	7.01	49.56
23 Manufacture of Other Non-metallic Mineral Products	7.73	7.50	6.65	52.63
24 Manufacture of Basic Metal Products	11.83	24.84	10.76	139.96
25 Manufacture of Fabricated Metal Products, Except Machinery and Furniture	4.07	3.75	5.43	24.22
26 Manufacture of Electronic Components, Computer, Radio, Television and Communication Equipment and Apparatuses	22.60	28.62	24.04	391.02
27 Manufacture of Medical, Precision and Optical Instruments, Watches and Clocks	5.39	6.98	9.36	28.45
28 Manufacture of electrical equipment	6.54	6.59	8.98	58.16
29 Manufacture of Other Machinery and Equipment	5.16	4.09	5.38	38.73
30 Manufacture of Motor Vehicles, Trailers and Semitrailers	25.23	22.87	51.78	399.99
31 Manufacture of Other Transport Equipment	25.19	14.25	10.46	569.10
32 Manufacture of Furniture	5.65	6.69	6.89	18.04
33 Other manufacturing	5.57	6.06	7.81	9.84
D. Electricity, gas, steam and water supply	9.25	12.77	23.30	117.83
E. Sewerage, waste management, materials recovery and remediation activities	6.47	7.77	5.75	19.86
F. Construction	4.41	3.86	4.23	52.48
G. Wholesale and retail trade	2.64	2.16	2.24	10.21
H. Transportation	3.75	11.91	4.27	30.48
I. Accommodation and food service activities	2.84	2.48	2.29	7.06
J. Information and communications	10.85	7.62	5.03	81.12
K. Financial and insurance activities	5.57	3.90	5.30	66.62
L. Real estate activities and renting and leasing	4.06	3.92	3.74	13.01
M. Professional, scientific and technical activities	4.41	7.66	5.94	74.67
N. Business facilities management and business support services	5.42	7.40	7.16	96.75
P. Education	2.94	4.54	5.20	42.24
Q. Human health and social work activities	5.30	4.04	4.79	60.24
R. Arts, sports and recreation related services	4.75	6.80	3.78	18.18
S. Membership organizations, repair and other personal services	3.44	3.26	2.66	5.64
Total	1.82	1.91	3.28	328.20

* NA : denotes census stratum.

Table 3.2 Distributions of the population (Pop.) and the sample (Sam.) sizes

KSIC9 Section and Division Code	Number of workers								Total	
	1~9		10~49		50~299		300~		Pop.	Sam.
	Pop.	Sam.	Pop.	Sam.	Pop.	Sam.	Pop.	Sam.		
A. Agriculture, forestry and fishing	1470	32	787	23	96	9	1	1	2354	65
B. Mining and quarrying	1394	30	349	16	22	5	5	3	1770	54
C. Manufacturing	264343	1492	52875	730	8994	372	601	144	326813	2738
10 Manufacture of Food Products	49078	137	3288	40	699	22	34	7	53099	206
11 Manufacture of Beverages	922	27	183	13	60	9	5	3	1170	52
12 Manufacture of Tobacco Products	1	1	3	2	8	3	1	1	13	7
13 Manufacture of Textiles, Except Apparel	16352	87	2791	38	378	17	7	3	19528	145
14 Manufacture of wearing apparel, Clothing Accessories and Fur Articles	19672	95	2505	36	268	15	25	5	22470	15
15 Tanning and Dressing of Leather, Manufacture of Luggage and Footwear	3548	45	683	21	69	9	2	1	4302	76
16 Manufacture of Wood Products of Wood and Cork ; Except Furniture	5212	57	748	22	41	7	2	2	6003	88
17 Manufacture of Pulp, Paper and Paper Products	3897	46	1325	28	209	14	11	3	5442	91
18 Printing and Reproduction of Recorded Media	14909	86	1283	27	92	10	0	0	16284	123
19 Manufacture of Coke, hard-coal and lignite fuel briquettes and Refined Petroleum Products	142	11	127	11	12	4	5	3	286	29
20 Manufacture of chemicals and chemical products except pharmaceuticals, medicinal chemicals	4833	50	2009	33	419	19	39	6	7300	108
21 Manufacture of Pharmaceuticals, Medicinal Chemicals and Botanical Products	409	20	251	14	179	13	8	2	847	49
22 Manufacture of Rubber and Plastic Products	11671	71	4457	45	704	23	31	8	16863	147
23 Manufacture of Other Non-metallic Mineral Products	6843	61	2156	34	258	15	19	8	9276	118
24 Manufacture of Basic Metal Products	3821	46	2134	34	457	19	39	11	6451	110
25 Manufacture of Fabricated Metal Products, Except Machinery and Furniture	45731	135	7999	57	838	24	24	7	54592	223
26 Manufacture of Electronic Components, Computer, Radio, Television and Communication Equipment and Apparatuses	5671	53	2966	39	854	24	132	18	9623	134
27 Manufacture of Medical, Precision and Optical Instruments, Watches and Clocks	5857	55	1723	31	257	15	8	4	7845	105
28 Manufacture of electrical equipment	12607	79	3281	40	549	21	44	10	16481	150
29 Manufacture of Other Machinery and Equipment	24171	100	7521	56	1051	26	52	10	32795	192
30 Manufacture of Motor Vehicles, Trailers and Semitrailers	3554	45	2459	36	931	25	78	17	7022	123
31 Manufacture of Other Transport Equipment	1150	28	919	24	485	18	26	10	2580	80
32 Manufacture of Furniture	8998	68	1189	26	99	11	7	3	10293	108
33 Other manufacturing	15294	89	875	23	77	9	2	2	16248	123
D. Electricity, gas, steam and water supply	737	25	427	16	304	16	31	7	1499	64
E. Sewerage, waste management, materials recovery and remediation activities	3388	43	1800	31	213	14	1	1	5402	89
F. Construction	72968	163	20248	83	3271	42	346	20	96833	308
G. Wholesale and retail trade	845359	423	28576	95	2571	37	148	13	876654	568
H. Transportation	335934	303	8305	58	2779	38	161	20	347179	419
I. Accommodation and food service activities	620926	353	12778	70	760	24	36	7	634500	454
J. Information and communications	18391	88	6248	52	1574	31	162	17	26375	188
K. Financial and insurance activities	20005	89	17287	80	1912	33	149	15	39353	217
L. Real estate activities and renting and leasing	117969	206	7625	55	460	19	27	7	126081	287
M. Professional, scientific and technical activities	58990	141	9605	63	1771	33	235	20	70601	257
N. Business facilities management and business support services	26603	108	6359	52	2503	38	445	22	35910	220
P. Education	142351	218	16433	77	6958	48	222	17	165964	360
Q. Human health and social work activities	87242	145	17082	80	2508	36	180	17	107012	278
R. Arts, sports and recreation related services	100007	174	2332	35	574	20	35	9	102948	238
S. Membership organizations, repair and other personal services	366259	314	9269	61	752	22	13	3	376293	400
Total	3084336	4347	218385	1677	38022	837	2798	343	3343541	7204

4. 추정

제안된 표본설계 하에서 실시될 이러닝사업체조사 자료로부터 조사항목에 대한 업종별·종사자규모별 추정값과 세부항목인 업종별·종사자규모별 추정값 계산에 사용될 추정식을 얻고자 한다. 먼저 추정식에 사용하게 될 기본적인 기호인 첨자 ijk 는 업종층 i 의 j 번째 종사자규모층 내 k 번째 사업체를 나타낸다. 사업체실태조사를 위해 고려할 수 있는 표본가중치 w_{ijk} 는 다음과 같이 세 가지 요소들로 구성된다. 즉, 표본가중치 w_{ijk} 는 설계가중치 (w_{ijk}^D)에 무응답조정과 보정조정 (calibration adjustment)을 위한 계수가 각각 곱해져 다음의 식으로 계산된다.

$$w_{ijk} = w_{ijk}^D \times f_{ijk}^{NRF} \times f_{ijk}^{CAL} \quad (4.1)$$

Table 4.1에 나타난 설계가중치는 각 업체가 갖는 표본추출률의 역수 ($\pi_{ijk}^{-1} = N_{ij}/n_{ij}$)로 정의되며, 무응답조정은 표본업체 중 무응답업체가 갖는 설계가중치 손실분에 대한 조정인자 f_{hik}^{NRF} 를 마지막으로 보정조정법에 사용되는 조정계수는 f_{hik}^{CAL} 이다. 여기서 Table 4.1의 설계가중치는 이러닝사업체조사통계의 시계열 유지를 위해, 나머지 두 조정인자들과는 달리, 매 조사마다 동일한 값으로 적용된다. 또한 기존의 단순집계를 벗어나 업종 및 종사자 규모별, 업종별 그리고 종사자 규모별 추정과 추정의 정도에 대한 평가에도 설계가중치가 유용하게 사용된다. 무응답조정은 세부 규모층별 표본대체를 실시하여 무응답을 일부 조정하지만, 전수조사층에서 무응답이 발생한 경우는 대체표본이 불가능하다. 이러한 경우는 설계가중치에 대한 단위무응답조정이 필요하다. 본 연구에서는 OECD (2003, pp. 73)에 근거하여, 설계가중치와 세부층내 무응답 조정을 고려하여 다음과 같이 가장 단순한 형태의 조정인자를 식 (4.1)에 사용하였다.

$$f_{ijk}^{NRF} = \frac{n_{ij}}{n_{ij}} \quad (4.2)$$

식 (4.2)에서 n_{ij} 는 (ij)층에서 응답한 조사단위 (표본 기업체)의 수를 나타낸다. 무응답에 대한 가중치 조정은 무응답에 따른 편향을 제거하는데 있지만 일반적으로 가중치 변동량이 증가하여 추정량의 추정오차도 증가하게 된다. 따라서 가중치 조정에 따른 편향축소와 추정오차 증가 간의 적절한 절충이 필요하다. 한편 특정변수 모총합에 대한 제한적인 정보만 갖는 경우, 보조변수에 대한 정보를 이용하여 해당 특성치의 모총합이 표본가중합과 일치하도록 하는 방법으로 가중치를 조정하는 방법을 보정조정법이라 부른다. 이 방법은 무응답에 따른 편향을 줄여 주는 효과와 표본틀 결함에 대한 보정에도 효과적이며, f_{hik}^{CAL} 는 라그랑지 승수법에 의해 구할 수 있다 (Deville과 Sarndal, 1992).

이러닝사업체조사 특정치인 이러닝지출액 y 의 모총합 $Y = \sum_{i=1}^{41} \sum_{j=1}^4 \sum_{k=1}^{n_{ij}} y_{ijk}$ 은 표본가중합 (weighted sample total)의 형태로 다음과 같이 계산된다.

$$\hat{Y} = \sum_{i=1}^{41} \sum_{j=1}^4 \sum_{k=1}^{n_{ij}} w_{ijk} y_{ijk} \quad (4.3)$$

단, w_{ijk} 는 식 (4.1)에서 $f_{ijk}^{CAL} = 1$ 인 가중치를 나타낸다. 총합추정량 \hat{Y} 을 층별 단순평균 $\bar{y}_i = \sum_j y_{ij}/n_i$, $n_i = \sum_j n_{ij}$, $\hat{Y} = \sum_i \bar{y}_i n_i$ 로 표현하면 층간 독립추출의 성질과 기초표본이론에 의하여 이러닝 기업체지출액 합 의 분산추정량은 다음과 같이 계산된다.

$$v(\hat{Y}) = \sum_{h=1}^{41} v(\bar{y}_i) = \sum_{i=1}^{41} \frac{s_i^2}{n_i} (1 - f_i) = \sum_{i=1}^{41} \frac{1 - f_i}{n_i(n_i - 1)} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (4.4)$$

Table 4.1 Design weight ($w_{ijk}^D=1/\text{sample proportion}$) for each stratum

KSIC9 Section and Division Code	Number of workers				Total
	1~9	10~49	50~299	300~	
A. Agriculture, forestry and fishing	0.022	0.029	0.094	1.000	0.028
B. Mining and quarrying	0.022	0.046	0.227	0.600	0.031
C. Manufacturing	0.006	0.014	0.041	0.240	0.008
10 Manufacture of Food Products	0.003	0.012	0.031	0.206	0.004
11 Manufacture of Beverages	0.029	0.071	0.150	0.600	0.044
12 Manufacture of Tobacco Products	1.000	0.667	0.375	1.000	0.538
13 Manufacture of Textiles, Except Apparel	0.005	0.014	0.045	0.429	0.007
14 Manufacture of wearing apparel, Clothing Accessories and Fur Articles	0.005	0.014	0.056	0.200	0.007
15 Tanning and Dressing of Leather, Manufacture of Luggage and Footwear	0.013	0.031	0.130	0.500	0.018
16 Manufacture of Wood Products of Wood and Cork ; Except Furniture	0.011	0.029	0.171	1.000	0.015
17 Manufacture of Pulp, Paper and Paper Products	0.012	0.021	0.067	0.273	0.017
18 Printing and Reproduction of Recorded Media	0.006	0.021	0.109	0.000	0.008
19 Manufacture of Coke, hard-coal and lignite fuel briquettes and Refined Petroleum Products	0.077	0.087	0.333	0.600	0.101
20 Manufacture of chemicals and chemical products except pharmaceuticals, medicinal chemicals	0.010	0.016	0.045	0.154	0.015
21 Manufacture of Pharmaceuticals, Medicinal Chemicals and Botanical Products	0.049	0.056	0.073	0.250	0.058
22 Manufacture of Rubber and Plastic Products	0.006	0.010	0.033	0.258	0.009
23 Manufacture of Other Non-metallic Mineral Products	0.009	0.016	0.058	0.421	0.013
24 Manufacture of Basic Metal Products	0.012	0.016	0.042	0.282	0.017
25 Manufacture of Fabricated Metal Products, Except Machinery and Furniture	0.003	0.007	0.029	0.292	0.004
26 Manufacture of Electronic Components, Computer, Radio, Television and Communication Equipment and Apparatuses	0.009	0.013	0.028	0.136	0.014
27 Manufacture of Medical, Precision and Optical Instruments, Watches and Clocks	0.009	0.018	0.058	0.500	0.013
28 Manufacture of electrical equipment	0.006	0.012	0.038	0.227	0.009
29 Manufacture of Other Machinery and Equipment	0.004	0.007	0.025	0.192	0.006
30 Manufacture of Motor Vehicles, Trailers and Semitrailers	0.013	0.015	0.027	0.218	0.018
31 Manufacture of Other Transport Equipment	0.024	0.026	0.037	0.385	0.031
32 Manufacture of Furniture	0.008	0.022	0.111	0.429	0.010
33 Other manufacturing	0.006	0.026	0.117	1.000	0.008
D. Electricity, gas, steam and water supply	0.034	0.037	0.053	0.226	0.043
E. Sewerage, waste management, materials recovery and remediation activities	0.013	0.017	0.066	1.000	0.016
F. Construction	0.002	0.004	0.013	0.058	0.003
G. Wholesale and retail trade	0.001	0.003	0.014	0.088	0.001
H. Transportation	0.001	0.007	0.014	0.124	0.001
I. Accommodation and food service activities	0.001	0.005	0.032	0.194	0.001
J. Information and communications	0.005	0.008	0.020	0.105	0.007
K. Financial and insurance activities	0.004	0.005	0.017	0.101	0.006
L. Real estate activities and renting and leasing	0.002	0.007	0.041	0.259	0.002
M. Professional, scientific and technical activities	0.002	0.007	0.019	0.085	0.004
N. Business facilities management and business support services	0.004	0.008	0.015	0.049	0.006
P. Education	0.002	0.005	0.007	0.077	0.002
Q. Human health and social work activities	0.002	0.005	0.014	0.094	0.003
R. Arts, sports and recreation related services	0.002	0.015	0.035	0.257	0.002
S. Membership organizations, repair and other personal services	0.001	0.007	0.029	0.231	0.001
Total	0.001	0.008	0.022	0.123	0.002

여기서 $y_{ij} = \sum_{k=1}^{n_{ij}} (n_i w_{ijk}) y_{ijk}$ 이고, $\bar{y}_i = \sum_j y_{ij} / n_i$ 는 조사변수 y 의 i 번째 업종 층의 모집단 총합인 $Y_i = \sum_{j=1}^4 \sum_{k=1}^{N_{ij}} y_{ijk}$ 의 불편추정량, $f_i = n_i / N_i$ ($N_i = \sum_{j=1}^4 N_{ij}$) 는 i 번째 업종 층의 표본추출률을 $s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ 는 i 번째 업종 층의 표본분산을 각각 나타낸다. 따라서 수요부문 기업체

이러닝 총 지출액 추정량의 표준오차와 변동계수는 다음과 같다.

$$SE(\hat{Y}) = \sqrt{v(\hat{Y})}, \quad CV(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}}.$$

참고로, 중업원 수에 따라 순서화 시킨 부모집단에서 추출한 계통표본내 조사단위들은 단순확률표본에 의해 추출한 조사단위에 비해 상대적으로 이러닝 지출액이 서로 차이가 큰 편으로 급내상관계수 (Kim 등, 2006)가 음수가 되는 경향이 있어 이러닝 지출액추정량의 분산을 과소추정하게 된다.

따라서 추정된 분산의 해석 및 검토가 필요하다. 그리고 무응답조정과 보정조정을 실시한 복합 표본조사의 경우, 식 (4.3)을 사용하여 모집단의 총합을 추정할 수 있으나, 총합추정량 \hat{Y} 의 분산식은 식 (4.4)와는 다른 형태를 가진다. 복합 표본조사시 필요한 \hat{Y} 의 분산식 유도에는 Park과 Whang (2012) 등을 참고 할 수 있다.

5. 결론

이러닝산업통계는 이러닝산업 전반에 걸친 수요와 공급을 망라한 광범위한 조사를 실시하고 있으며, 조사내용에서도 단체 수요자, 가구, 개인 및 공급자를 모두 포함하고 있어 본 통계는 활용도가 매우 높은 통계이다. 따라서 본 통계가 우리나라의 이러닝산업을 제대로 반영하고 있는지에 대한 포괄적인 점검이 필요하였다. 이에 대한 단기과제로서 이러닝사업체수요부문의 모집단의 재설계가 필요하였으며, 실태조사계획 전반에 걸쳐 개선이 요구되었다. 이에 본 연구는 2010년에 통계청이 발표한 전국사업체조사통계와 KSIC9를 기초자료로 활용하여 조사모집단을 업종별·종사자규모별로 층화하고, 락배정을 사용한 층화계통추출법에 의해 새로운 표본설계안을 제시하였다. 이를 위해 통계청 품질승인기준을 만족시키도록 목표오차를 정하고 이에 적절한 락배정식 승수값이 $p=0.4$ 임을 모의표본실험을 통해 찾아내었으며, 새로운 표본설계안 하에서의 가중치 산출과 모수추정식도 제시하였다. 따라서 본 연구의 결과는 현 단계에서 이러닝사업체조사의 문제를 보완하여 기업체실태조사통계의 품질개선 및 정확도의 향상에 이바지할 것이다. 한편 과년도 조사에 의하면 중요변수 (이러닝 지출액)에 대한 항목무응답비율이 높아 (3.8% 2010년 기준, 새로운 표본설계로 2013년 이러닝사업체조사를 실시할 경우, 핫덱 (hot deck)대체, 층 평균값 대체 등 대체방법 (Kim 등, 2006) 등 항목무응답 값을 대체하는 방법에 대한 검토 및 특이치 검출 및 처리방안 (Lee, 2012)과 함께 표본설계의 개정에 따른 사업체조사통계의 시계열 유지에 대한 연구도 추가로 필요할 것이다.

References

- Deville, J. and Sarndal, C. E. (1992). Calibration estimator in survey sampling. *Journal of the American Statistical Association*, **87**, 376-382.
- Heo, S. and Chang, D. J. (2010). A sample survey design for service satisfaction evaluation of regional education offices. *Journal of the Korean Data & Information Science Society*, **21**, 669-679.
- Kim, D. H. and Hwang, J. S. and Kwak, S. G. (2010). A sample design for the survey on actual state of SMEs. *Journal of the Korean Data & Information Science Society*, **21**, 1021-1029.
- Kim, Y. W. and Ryu, J. B. and Park, J. W. and Hong, G. H. (2006). *Elementary survey sampling*, 6th Ed., Tomson Korea Limited, Seoul.
- Korea National Statistics Office. (2008). *Korean standard industrial classification*, A e-book, Korea National Statistics Office, Seoul.
- Lee, H. J. and Kang, S. B. (2012). Handling the nonresponse in sample survey. *Journal of the Korean Data & Information Science Society*, **23**, 1183-1194.
- Lee, K. J. (2012). *S tudy on outlier detection and management in industry status study*, Economic Statistics Bureau, The Bank of Korea, Seoul.

- Ministry of Knowledge Economy-National IT industry Promotion Agency (2012). *2011 e-learning industry status survey*, National IT industry Promotion Agency, Seoul.
- Ministry of Knowledge Economy (2012). *Assessment of quality of 2011 e-learning industry status survey*, Statistics Korea, Ministry of Knowledge Economy, Daejeon.
- OECD. (2003). *Business tendency surveys: A handbook*, OECD, Paris.
- Park, I. H. and Whang, H. J. (2012). *Study on variance estimation of sampling survey statistics proceeded by the Bank of Korea*, Economic Statistics Bureau, The Bank of Korea, Seoul.
- Sohn, K. C. and Kim, D. H. (2012). Study for the sampling method using simulation in clinical data. *Journal of the Korean Data & Information Science Society*, **23**, 677-682.

A sampling design for e-learning industry status survey on the business demand sector

Hea-Jung Kim¹ · Hwa-Ryun Kwak²

¹Department of Statistics, Dongguk University

²Institute of Statistical Information and Technique, Dongguk University

Received 10 May 2013, revised 12 June 2013, accepted 20 June 2013

Abstract

The e-learning industry status survey statistic provides information about the actual conditions of supply and demand of the e-learning industries. NIPA (National IT Industry Promotion Agency) has published the annual report of the survey results since 2004. Due to the 9th version of the KSIC (Korean standard industrial classification) revised in 2008, a refinement of the sampling design for the survey becomes necessary, especially that for the business demand sector. This article, based on the 9th revision of the KSIC, constructs a stratification of the target population used for the e-learning industry status survey on the business demand sector. Classification of strata in the business population is based on the industrial type and employment scale of business. Under the stratified population, we design a sampling scheme by using the power allocation method that enables us to satisfy a target coefficient of variation of each industrial stratum. In order to secure an accurate survey results based on the proposed sampling design, we consider the problem of calculating the design weights, derivation of parameter estimators, and formulas of their standard errors.

Keywords: E-learning industry status survey statistic, Korean standard industrial classification, power allocation, sampling design, stratification.

¹ Corresponding author: Professor, Department of Statistics, Dongguk University, Seoul 100-715, Korea.
E-mail: kim3hj@dongguk.edu

² Researcher, Department of Statistics, Dongguk University, Seoul 100-715, Korea.