

# 노래연습장 저작권사용료 분배방법 개선을 위한 표본설계연구

이계오<sup>1,a</sup>, 최정배<sup>b</sup>

<sup>a</sup>한남대학교 비즈니스통계학과, <sup>b</sup>공군사관학교 기초과학과

## 요약

노래연습장에서 연주되는 노래곡목별 히팅수를 정확하게 추정하기 위한 표본설계에서는 조사모집단 정의와 층화방안을 연구하고 기존데이터를 분석하여 표본크기를 산출한 후에 표본분배방법을 연구하였다. 표본설계에 의해서 선정된 표본업소들의 대표성을 온라인 노래연습장의 로그데이터를 이용하여 모수추정치 오차를 분석하여 검토하였다. 오프라인 노래연습장의 모수추정법을 제안하고 실제로 수집한 로그데이터로부터 모수추정치를 계산하여 온라인과 오프라인 노래연습장에서 연주되는 노래곡목별 히팅수에서 차이가 있음을 보였다. 온라인과 오프라인 노래연습장의 특성차이를 반영한 분배방법의 개선안을 제안하고 비교분석을 통해서 최적 분배방법을 선정하였다.

주요어: 조사모집단 정의와 층화, 온라인과 오프라인 노래연습장, 모수추정법.

## 1. 서론

음악인의 창작활동을 장려하기 위해서 음악공연활동이나 음악 이용시에는 사용료를 징수하고 이를 저작권자에게 공정하게 분배하기 위해서 (사)한국음악저작권협회(KOMCA)가 설립되었다. 저작권 사용료 징수에 관련된 규정은 1988년 02월 23일 제정된 후에 14차례 변경을 거쳐서 현행규정은 2008년 02월 28일에 확정되었으며 이를 기준으로 징수하고 있다. 반면에, 음악저작물 사용료 분배규정은 1983년 02월 23일에 제정된 이래 6차례의 변경을 거쳐 현행규정을 2008년 09월 12일에 확정하여 관계권리자에게 분배 적용하고 있다. 그런데 현행 분배규정의 개정 배경에는 관계권리자들에게 적용하는 분배기준이 투명하지 못하고 일부 데이터들이 조작되는 문제점이 발견되었다. 이에 대한 정확한 실태를 파악하지 않고 주먹구구식 의사결정으로 기존 분배규정에 비해서 분배금액이 급격하게 감소된 관계권리자들의 강한거부와 관련부처 앞의 연이은 시위 등의 사회적인 문제가 야기되었다.

현행분배규정의 문제점을 정확하게 파악하고 이에 대한 합리적인 개선방안을 마련하기 위한 표본설계연구가 필요하게 되었다. 본 연구는 우선 노래방에서 불러지는 노래 곡목별 히팅수의 실태를 파악할 수 있는 표본설계에 관한 연구를 2장에서 다루고 3장에서는 표본설계에 의해서 추출된 노래연습장들이 실제 조사모집단을 대표하는지의 타당성을 검토하고 4장에서는 오프라인 노래연습장의 노래 곡목별 히팅수의 추정방법을 제안하고 실제 조사된 데이터를 이용한 추정결과를 분석한 후에 5장에서는 분배방법의 개선안을 제안하고 비교분석을 통해서 최선안을 선정한다. 마지막 결론에서는 연구내용을 요약하고 분배규정의 개정방안을 제안하고자 한다.

<sup>1</sup> 교신저자: (306-791) 대전광역시 대덕구 오정동 133번지 한남대학교 비즈니스통계학과, 교수.  
E-mail: kayolee@hnu.kr

## 2. 표본설계

### 2.1. 조사목적 및 모집단 정의

현행분배규정에서는 노래연습장의 업소형태에 관계없이 온라인기준으로 저작권료를 분배중이나 이에 대한 정확한 실태를 파악하여 현행제도의 타당성을 검토한 후, 개선방안을 제시하고자 하는데 이때 자료수집에 사용할 표본업소를 추출하는 방법을 연구하고자한다. 주요 연구내용은 노래연습장의 곡목별 정확한 연주실태(히팅수)를 근거로 하여 조사모집단을 대표할 수 있는 표본추출법을 연구하는 것이다.

모집단으로 사용할 데이터는 2009년 1월 기준 (사)한국음악저작권협회에서 관리중인 노래연습장의 명부에 등재되어 영업 중인 28,540개 업소를 조사모집단으로 정의한다.

### 2.2. 모집단 분석 및 층화

모집단 분석은 표본조사대상인 노래연습장 업소들의 특성을 파악하여 층을 동질적으로 구분하는데 필요한 층화변수를 탐색하기 위한 분석이다.

우선 조사모집단을 시도별, 시도규모별, 업소형태별로 분류하여 업소수와 반주기수의 빈도를 정리하면 다음 표 1과 같다.

표 1에서 보면 노래연습장에서 온라인형태는 3,072개로 10.76%를 차지하고 있으며 서울시, 인천시와 경기도에 편중되어 있음을 볼 수 있다. 반면에 오프라인 형태는 24,184개로 경기도와 서울이 다수를 차지하고 있으나 비울로는 경기도를 제외한 도지역이 높은 구성비를 갖고 있으며 혼합형은 온라인형태와 유사한 특성을 보이고 있다. 또한 시도규모와 업소형태에 대한 교차빈도표를 이용하여 분포적 특성을 분석해보면 중소도시에서 업소당 기기수는 온라인은 8.779대, 혼합은 7.557대이고 오프라인은 5.452대로 온라인 업소의 규모가 큰 것으로 나타났다.

다음으로 노래연습장의 규모에 대한 구분은 기기 5대씩을 기준으로 '1': 1-5대, '2': 6-10대, '3': 11-15대, '4': 16-20대, '5': 21-25대, '6': 26-30대, '7': 31대 이상과 같이 분류하면 분포특성은 규모1이 전체의 53%정도, 규모2가 전체의 43.5%를 차지하여 두 개의 층에 집중적으로 분포되어 있으며, 온라인으로 영업하는 노래연습장은 규모가 클수록 상대적인 구성비가 크게 나타나는 특성을 보이고 규모3 이상의 경우에는 오프라인의 업소보다 더 많은 업소를 포함하는 특성을 볼 수 있으므로 규모가 클수록 온라인 형태로 영업하는 노래연습장이 많다는 의미로 해석할 수 있을 것이다.

주요 조사내용인 연주 실태에 영향을 미치는 요인을 파악하기 위해서 기존의 온라인반주기와 네트워크 반주기에서 수집한 데이터를 분석한 후에 핵심변인을 찾아내고 이를 층화변수로 사용하여 모집단을 동일한 특성을 갖는 층으로 세분화하는 것을 모집단의 층화(stratification)라고 한다 (박홍래, 2000).

동일한 층에 속하는 노래연습장들은 유사한 특성을 갖기 때문에 소수의 표본을 추출하여도 정확한 추정치를 산출할 수 있으므로 사회실태조사와 사업체 실태조사에서 주로 사용하고 있는 모집단을 층화하여 표본을 추출하는 층화추출법을 사용하고자 하며 노래연습장의 표층화변수로는 시도, 시도규모, 업소형태, 업소규모를 고려한 층화추출법을 제시하고자 한다.

### 2.3. 표본크기 검토 및 표본배분

#### 2.3.1. 산출 공식

표본크기는 표본추출법과 신뢰계수 및 모집단의 특성의 영향을 받기 때문에 먼저 단순임의 추출법에서 표본크기를 산출하는 공식을 살펴보고 층화추출법에서 공식을 구하자. 추정하고자하는 것이 모평균인 경우에 모집단의 크기가  $N$ 이고 여기서 표본크기를  $n$ 으로 가정한다. 표본은 단순임의추출법으로

표 1: 시도별, 시도규모별, 업소형태별 노래연습장의 업소수와 반주기수

시도	중소도시								군지역								합계	
	온라인		오프라인		혼합		합계		온라인		오프라인		혼합		합계		업소	기기
	업소	기기	업소	기기	업소	기기	업소	기기	업소	기기	업소	기기	업소	기기	업소	기기		
강원	117	919	531	2,891	41	294	689	4,104	36	284	147	938	11	93	194	1,315	883	5,419
경기	721	6,553	4,853	27,293	261	1,935	5,835	35,781	22	144	141	772	13	82	176	998	6,011	36,779
경남	129	1,018	1,433	7,424	36	263	1,598	8,705	6	33	156	790	5	31	167	854	1,765	9,559
경북	120	943	1,191	6,012	54	355	1,365	7,310	13	91	238	1,208	12	63	263	1,362	1,628	8,672
광주	52	499	903	4,644	21	176	976	5,319	.	.	.	.	.	.	0	0	976	5,319
대구	89	725	1,742	8,730	61	436	1,892	9,891	.	.	.	.	.	.	0	0	1,892	9,891
대전	116	1,017	949	4,826	41	336	1,106	6,179	.	.	.	.	.	.	0	0	1,106	6,179
부산	152	1,546	1,965	10,361	71	564	2,188	12,471	.	.	.	.	.	.	0	0	2,188	12,471
서울	675	6,228	4,141	23,977	357	2,835	5,173	33,040	.	.	.	.	.	.	0	0	5,173	33,040
울산	61	514	808	3,916	16	102	885	4,532	.	.	.	.	.	.	0	0	885	4,532
인천	275	2,176	1,477	8,452	108	721	1,860	11,349	.	.	.	.	.	.	0	0	1,860	11,349
전남	79	549	510	2,425	26	184	615	3,158	14	106	197	993	8	58	219	1,157	834	4,315
전북	91	868	644	3,817	29	266	764	4,951	8	72	87	474	1	9	96	555	860	5,506
제주	50	421	229	1,389	36	275	315	2,085	.	.	.	.	.	.	0	0	315	2,085
충남	80	709	689	3,756	24	204	793	4,669	48	347	297	1,568	11	76	356	1,991	1,149	6,660
충북	97	811	620	3,768	30	213	747	4,792	21	158	236	1,318	11	79	268	1,555	1,015	6,347
합계	2,904	25,496	22,685	123,681	1,212	9,159	26,801	158,336	168	1,235	1,499	8,061	72	491	1,739	9,787	28,540	168,123

로 추출하는데 모평균에 대한 추정치와 참값간의 허용오차를  $B$ 로 가정하고 신뢰계수를  $z_{\alpha/2}$ 라고 한다면 허용오차는 다음 식으로 표현된다 (이계오 등, 2007).

$$B = z_{\frac{\alpha}{2}} \sqrt{\frac{N-n}{N} \frac{S^2}{n}}, \tag{2.1}$$

여기서,  $S^2$ 은 모집단의 분포특성인 분산을 나타낸다. 식 (2.1)에서 양변을 제곱하고  $n$ 에 관해서 정리하면 표본크기 산출 식은 아래와 같다.

$$n = \frac{\left(z_{\frac{\alpha}{2}} S\right)^2}{B^2 + \frac{1}{N} \left(z_{\frac{\alpha}{2}} S\right)^2} = \frac{n_0}{1 + \frac{n_0}{N}}, \quad \text{단 } n_0 = \frac{\left(z_{\frac{\alpha}{2}} S\right)^2}{B^2}. \tag{2.2}$$

노래연습장연구에서는 히팅수에 영향을 주는 요인이 시도, 시도규모, 업소형태와 업소규모이므로 이들을 층화변수로 하여 모집단을 층화하였다. 우선 전체 표본크기는  $n$ 으로 가정하고 층별 표본배분은 산출식의 도출의 편의를 위해서 비례할당법을 적용한 것으로 한다. 전체의 히팅수를 추정하는 것이 목적이지만 만일에 업소당 평균 히팅수를 추정한다면 전체 업소수를 평균 히팅수에 곱하면 전체 히팅수를 추정할 수 있으므로 평균 히팅수 추정에 대한 표본크기 산출 식을 도출 하겠다.

층화추출법은 각 층별로 표본을 독립적으로 추출하므로 층별 평균을 계산하고 모든 층에 대해서 층별 구성비를 기준으로 평균으로 계산되므로 모평균과 추정량을 아래 식으로 표현할 수 있다.

$$\mu = \sum_{h=1}^L W_h \mu_h, \quad \hat{\mu} = \sum_{h=1}^L W_h \hat{\mu}_h,$$

단,  $\hat{\mu}_h$ 는  $h$ 층의 표본평균이다. 위의 식을 이용하여 전체 히팅수를 추정하고자할 경우에는 전체 업소수  $N$ 을 평균 히팅수에 곱하면 되므로 아래와 같이 표현할 수 있다.

$$\hat{\tau} = N \cdot \hat{\mu} = N \cdot \sum_{h=1}^L W_h \hat{\mu}_h = \sum_{h=1}^L N_h \hat{\mu}_h, \tag{2.3}$$

단,  $W_h = N_h/N$ 이다.

식 (2.3)으로 주어진 추정량의 분산은 평균 히팅수의 추정량의 분산에  $N^2$ 을 곱해서 계산할 수 있으므로 아래와 같이 나타낼 수 있다.

$$V(\hat{\tau}) = N^2 V(\hat{\mu}) = \sum_{h=1}^L N_h^2 V(\hat{\mu}_h) = \sum_{h=1}^L N_h^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}. \quad (2.4)$$

전체 히팅수의 추정은 평균히팅수의 추정으로 계산하므로 앞으로는 모평균의 추정량에 대한 분산 추정을 살펴보자.

$$\hat{V}(\hat{\mu}) = \sum_{h=1}^L W_h^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}. \quad (2.5)$$

평균 히팅수의 추정치에 대한 오차의 한계가  $B$ 이하가 되도록 하는 표본크기의 결정은 신뢰계수가  $z_{\alpha/2}$ 라고 가정하면 아래와 같이 산출한다.

$$B = z_{\frac{\alpha}{2}} \sqrt{V(\hat{\mu})} = z_{\frac{\alpha}{2}} \sqrt{\sum_{h=1}^L W_h^2 \frac{N_h - n_h}{N_h} \frac{S_h^2}{n_h}}, \quad (2.6)$$

여기서  $V(\hat{\mu})$ 는 각 층의 표본수  $n_h$ 들의 식으로 표현되어 있다. 따라서 주어진 식을 전체 표본크기( $n$ )에 관하여 풀기 위해서는  $n$ 과  $n_h$ 의 관계식(각 층에 대한 표본 배분 비율)이 주어져야 한다. 전체 표본크기  $n$ 을 각 층에 어떻게 배분하는가의 문제는 층화임의추출법에서 대단히 중요한 문제이지만 여기서는 비례분법을 적용한다고 가정한다.

식 (2.6)에서 좌우 변을 제곱한 후에  $n$ 에 관해서 정리하면 전체 표본크기를 산출하는 식을 얻게 된다.

$$n = \frac{\sum_{h=1}^L W_h S_h^2}{V_0 + \frac{1}{N} \sum_{h=1}^L W_h S_h^2}, \quad (2.7)$$

여기서  $V_0$ 는 식 (2.6)을 이용하여  $(B/z_{\alpha/2})^2$ 으로 계산되며 적당한 허용오차를 선정해가면서 조사수행에 문제가 없을 정도의 표본크기를 계산하면 된다.

### 2.3.2. 표본크기 검토

전체 노래연습장업소(28,540개)에 대한 업소 당 평균 히팅수의 허용오차를 500곡으로 할지라도 전체 히팅수에 대한 허용오차는 14,270,000곡으로 이것이 월간의 전체 히팅곡 수 추정에서 실질적인 의미를 갖게 될 수 없기 때문에 허용오차를 축소하면서 표본크기를 산출하였다. 다양한 허용오차에 대한 표본크기의 계산결과를 아래 표 2에 요약하였다.

표 2에 주어진 여러 가지 표본크기 중에서 추출률이 5%정도이고 업소 당 평균 히팅수의 오차가 395곡인이며 총 히팅수의 허용오차는 12,280,000곡 이내인 1,500개와 추출률이 4.38%이고 총 히팅수의 허용오차가 12,413,300곡 이내인 1,250개를 중심으로 표본업소의 로그데이터 취합기의 설치비, 표본업소 관리 및 데이터 분석 등에 사용되는 비용, 허용오차 등을 종합적으로 고려하고 문화관광체육부 담당공무원, 한국음악저작권협회 관계자와 외부전문가들의 자문을 거쳐서 최종적으로 표본크기는 1,250개로 확정하였다. 그러나 상황변동에 대비하여 표본크기 1,500개 방안에 대해서도 표본배분법을 동시에 제시한다.

표 2: 노래연습장 허용오차와 표본크기

평균 허용오차	총계 허용오차	표본크기	표본추출률
700	19,978,000	496	0.0174
600	17,124,000	671	0.0235
488.4370	13,940,000	1,000	0.0350
454.2867	12,765,400	1,150	0.0403
434.9439	12,413,300	1,250	0.0438
420.4625	12,000,000	1,334	0.0467
395.2340	11,280,000	1500	0.0526
300	8,562,000	2,507	0.0878
250	7,135,000	3,475	0.1218
200	5,708,000	5,082	0.1781

### 2.3.3. 표본배분

표본크기 계산에서 표본배분은 비례배분을 가정하였으나 비례배분의 기준을 층별로 업소수로 하는 방안과 층별로 반주기수로 하는 방안을 함께 검토하였다.

층의 구분은 표본크기를 산출할 때 사용한 층화변수인 시도, 업소형태(온라인, 오프라인, 혼합)와 업소규모를 기준으로 하였으며 층별 배분을 비례할당법을 적용하는데 기준변수로는 업소수와 반주기수를 사용하였다. 표본크기 산출에 온라인업소에서만 축적한 로그데이터만을 이용하였으나 온라인과 오프라인의 층의 표준편차가 유사하다는 가정에서는 표본크기 계산결과는 타당성을 가질 것이다. 만일에 각 층별로 표준편차가 오프라인과 온라인의 업소간의 특성에서 차이가 없다면 표본배분법으로 내이만배분법을 적용하는 것이 적절하겠지만 또 다른 층화요인인 제작사를 반영하기 위해서는 비례배분법이 로버스트(robust)할 것으로 판단하였기 때문에 비례배분법만을 고려하였다.

우선 노래연습장의 표본배분에서는 업소수와 반주기수를 기준으로 비례배분을 층별로 계산하였으며 여기서 층별로 최소모집단의 규모를 30으로 제한하였기 때문에 층을 통합하기 위해서 층화변수의 우선순위는 시도, 업소형태(온라인, 오프라인, 혼합), 시도규모와 업소규모의 순서로 하였다. 우선 업소수를 기준으로 표본배분 결과는 아래 표 3에 정리하였다.

표 3에서 업소형태별로 집계해보면 1(온라인)은 135업소, 2(오프라인)은 1,060업소, 3(온라인)은 55업소로 할당되는데 2(오프라인)가 약 95%를 차지하므로 추후 실제 자료수집 및 분석시에 필히 고려되어야 할 사항이다. 다음으로 반주기수를 기준으로 표본배분의 결과는 업소수를 기준으로 표본배분 결과와 대동소이하므로 반주기수를 기준으로 표본배분 결과를 확정한다.

### 2.4. 표본추출 및 자료수집방법

표본추출을 위해서 조사모집단을 층화변수인 시도별, 업소형태, 도시규모, 업소규모 등을 기준으로 정렬하고 각 층별로 최소한 30개 이상의 업소가 분포되도록 층을 통합한다. 통합된 모든 층에 대해서 조사모집단의 업소와 할당된 표본업소수를 정리한 후에 이들을 업소규모와 지역적인 주소를 기준으로 정렬한 다음에 추출간격에 따라서 계통추출법으로 표본업소를 선정한다.

노래연습장에서 층화변수로 고려되지 않은 제작사(태진과 금영)에 대해서는 표 3에 주어진 층별 할당결과를 각 층별로 제작사의 상대빈도를 기준으로 비례 배분한다.

추출된 표본 업소 내에서 어느 기기에 로그데이터 취합기를 부착할 것인지는 노래연습장의 기기수를 기준으로 계통추출법으로 기기를 선정하여 취합기를 부착하고 일정한 기간(예를 들어 분기단위)을 기준으로 부착기기를 순환하는 방식으로 관리할 것을 제안한다. 표본업소 추출의 정확성과 타당성을

표 3: 노래연습장 업소수 기준 표본배분

시도	업소수 (온/오프라인)	표본배분(1,250) (온/오프라인)	표본배분(1,500) (온/오프라인)
강원	153/678/52	7/30/2	8/36/3
경기	743/4,993/274	32/219/12	39/262/14
경남	135/1,589/41	6/70/2	7/84/2
경북	133/1,429/66	6/63/3	7/75/3
광주	52/903/21	2/39/1	3/47/1
대구	89/1,742/61	4/76/3	5/92/3
대전	116/949/41	5/41/2	6/50/2
부산	152/1,965/71	7/86/3	8/103/4
서울	675/4,141/357	29/182/16	36/207/19
울산	61/808/16	3/35/1	3/42/1
인천	275/1,477/108	12/64/5	14/78/6
전남	93/707/34	4/31/1	5/37/2
전북	99/731/30	4/32/1	5/38/2
제주	50/229/36	2/10/2	3/12/2
충남	128/986/35	6/43/2	7/52/2
충북	118/856/41	5/37/2	6/45/2
계	28,540	1,250	1,500

확보하기 위해서는 현재 저작권 사용료를 징수하는 업소들의 명부를 확보해야하고 이들의 업소형태와 규모를 정확하게 알아야 표본의 대표성을 제고할 수 있고 또한 히팅수 추정에서도 오차가 적어질 수 있다. 물론 표본업소가 선정되면 현재 모집단 자료에 있는 정보(반주기 수, 온라인/오프라인여부, 제작사 등)를 확인하여 모집단 데이터의 정확성을 간접적으로 확인할 수 있지만 전체 업소에 대한 마스터 파일의 정리와 관리는 필요한 작업이다.

표본추출당시에 사용한 조사모집단에 포함된 표본업소일지라도 조사당시에 현장을 방문하면 휴·폐업중이거나 본 조사에 참여를 거부할 경우에는 동일한 층내에서 특성이 유사한 업소로 대체해야할 것이다. 대체할 업소의 명부를 사전에 선정하여 예비표본업소로 관리하는 방법과 교체해야할 업소가 생길 때 마다 유사한 특성의 업소를 선정하는 방법이 있으나 조사현장에서 표본대체 절차만 잘 준수될 수 있다면 후자의 경우가 표본의 대표성 유지에 더 적합한 것으로 판단된다.

표본업소의 휴·폐업이나 응답거부 등으로 인한 표본업소교체 등이 있을 경우에는 이에 대한 정보를 별도로 기록 유지하여야하며 가중치 계산이나 모수추정에서 이 정보를 이용하게 된다. 만일에 새로 창업하는 업소에 대한 정보도 별도로 기록 유지하여 조사모집단의 변동을 가중치 산출에 반영해야한다.

선정된 표본업소에 대해서는 대외적으로 기밀유지가 중요하다. 혹시나 노출되어 로그데이터를 왜곡시킬 수 있는 상황이 발생한다면 해당업소의 로그데이터는 삭제하여야하고 즉시 표본업소를 교체하는 절차를 명시한 표본업소관리 지침 등이 마련되어야한다.

### 3. 표본설계의 타당성 분석

#### 3.1. 배경

표본설계연구에서는 오프라인의 노래연습장뿐 아니라 온라인 노래연습장도 오프라인 노래연습장과 동일한 방법으로 표본업소를 선정하였으므로 표본설계연구의 타당성을 검토하고 모수추정의 적합성을 검증하기위해서 온라인 노래연습장에서 수집한 로그데이터를 분석하겠다. 온라인 노래연습장의 표본업소가 전체 온라인 노래연습장을 대표할 수 있는 지를 파악하기 위해서 온라인 표본노래연습장에서 수집한 로그데이터로부터 가중치를 적용하여 추정한 히팅수와 전체 온라인노래연습장에서 계산한

곡목별 히팅수 간에 차이를 비교하여 통계적으로 유의한가를 검토하였다. 본 연구에서는 수치적인 검토를 위해서 2009년 2월의 온라인 노래연습장의 로그데이터를 이용하여 아래와 같은 절차로 곡목별 히팅수를 산출하였다.

### 3.2. 자료수집

온라인 노래연습장의 자료수집 절차는 온라인노래연습장 전체를 지부별(14개), 반주기제작사별(상위2개: 금영, 태진)로 분류하여 업소 수(조사모집단)와 전체 히팅수를 조사한 후에 표본설계연구에서 선정된 표본업소에서 랜덤하게 반주기를 선정하고 지부별, 제작사별로 분류된 업소수와 표본 업소수의 관계를 이용하여 가중치를 계산한다. 표본업소의 표본반주기에서 수집된 곡목별 히팅수에 계산된 가중치를 적용하여 곡목별 히팅수를 추정하고 이것과 전체 온라인 노래연습장의 곡목별 히팅수 비교분석한다.

### 3.3. 가중치 산출

표본노래연습장에서 반주기 하나를 표본으로 선정하였으므로 표본추출과정은 노래연습장을 1차로 추출하고 선정된 표본업소에서 하나의 반주기를 2차로 추출하였으므로 2단계추출과정에 의해서 표본반주기를 선정한 것으로 간주할 수 있다. 종합적으로 보면 각층별로 모집단의 반주기 수에서 표본업소 수만큼의 반주기(표본 반주기 수와 동일함)를 추출한 것으로 볼 수 있으므로 가중치(과정\_1)에서는 모집단 반주기 수를 표본업소수로 나누어 계산한다. 표본업소가 하나도 없으면 가중치의 합계와 모집단 반주기 수가 일치하지 않을 수 있기 때문에 이를 보정하는 것이 가중치(과정\_2)이다. 가중치(과정\_2)는 가중치(과정\_1)에 표본업소수를 곱하여 합계를 계산한 후에 표본가중치합계와 모집단의 반주기 수를 일치시키는 가중치를 벤치마킹보정가중치라고하며 이를 최종가중치로 사용한다 (Biemer와 Lyberg, 2003).

위의 과정결과에 대한  $i$ 지부에서  $j$ 제작사의 최종가중치( $W_{ij}$ )를 수식으로 표현하면 다음과 같다.

$$W_{ij} = \frac{M_{ij}}{n_{ij}} \times \frac{\sum_i^{14} \sum_{j=1}^2 M_{ij}}{\sum_i^{14} \sum_{j=1}^2 \frac{M_{ij}}{n_{ij}}}, \quad (3.1)$$

여기서,  $N_{ij}$ :  $i$ 지부에서  $j$ 제작사의 반주기가 설치된 조사모집단의 업소수,

$n_{ij}$ :  $i$ 지부에서  $j$ 제작사의 반주기가 설치된 표본 업소수,

$M_{ij}$ :  $i$ 지부에 설치된  $j$ 제작사의 전체 반주기 수,

$m_{ij}$ :  $i$ 지부의 표본업소에 설치된  $j$ 제작사의 반주기 수

이며,  $n_{ij}$ 가 0이면  $M_{ij}/n_{ij} = 0$ 으로 계산한다.

식 (3.1)을 이용하여 2009년 2월 한 달간 수집한 자료에 대해 지부별 제작사별로 가중치를 산출하면 다음 표 4와 같다.

### 3.4. 모수추정

표본업소에 대한 가중치와 관찰값이 주어진 경우에는 전체 히팅수와 특정 곡목별 히팅수에 대한 추정식은 모 총계에 관한 추정식과 같은 형태를 갖게 될 것이다 (Cochran, 1977).

아래 식은  $i$ 지부의  $j$ 제작사 반주기에서 곡목명  $l$ 의 히팅수 추정 값을 계산하기 위한 것이다.

$$\hat{t}_{ijl} = \sum_{i=1}^{14} \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} W_{lij} \times y_{ijkl}, \quad (3.2)$$

표 4: 온라인 노래연습장의 가중치 산출 결과

지부	제작사	모집단		표본		최종가중치
		업소수	반주기수	업소수	반주기수	
강남	금영/태진	146/102	1,229/876	4/4	36/36	312.7746/222.9378
강북	금영/태진	207/397	1,623/3,699	7/10	42/80	236.0261/376.5511
강서	금영/태진	94/140	796/1,215	3/5	18/40	270.1042/247.3693
강원	금영/태진	65/94	499/751	3/4	17/29	169.3241/191.1259
경기	금영/태진	488/325	3,723/2,830	16/10	120/84	236.8714/288.0885
경남	금영/태진	208/7	1,464/47	9/0	74/0	165.5915/0
경북	금영/태진	398/24	2,593/182	10/0	55/0	263.9624/0
부산	금영/태진	179/22	1,621/184	4/3	32/30	412.5367/62.43615
인천	금영/태진	13/333	95/2,525	0/13	0/88	0/197.7232
전남	금영/태진	179/10	1,177/97	6/0	34/0	199.6939/0
전북	금영/태진	83/13	720/156	4/4	29/0	183.2365/0
제주	금영/태진	22/35	160/308	1/1	6/7	162.8769/313.538
충남	금영/태진	227/97	1,662/906	8/8	62/19	211.4855/307.4302
충북	금영/태진	103/95	768/761	3/2	23/20	260.6031/387.3416
합계		4,107	32,667	133	981	

여기서,  $W_{ij}$ 는 식 (3.1)에서 주어진 가중치를 나타내고,  $y_{ijkl}$ 는  $(i, j)$ 층내의  $k$ 표본업소에서 수집한  $l$ 곡목의 히팅수를 나타낸다. 식 (3.2)에 주어진 추정량에 대한 분산 추정식은 일종의 층화비례추출법을 적용한 표본추출이므로 근사적으로 아래 식으로 계산할 수 있다.

$$\hat{V}(\tau_{ijl}) = \sum_{i=1}^{14} \sum_{j=1}^2 w_{ij}^2 \left( \frac{1}{n_{ij}} - \frac{1}{N_{ij}} \right) s_{ijl}^2, \quad (3.3)$$

여기서,  $w_{ij} = N_{ij} / \sum_{i=1}^{14} \sum_{j=1}^2 N_{ij}$ 는  $i, j$ 층의 조사모집단에서 구성비를 나타내고  $N_{ij}$ 와  $n_{ij}$ 는  $(i, j)$ 층에서 각각 모집단 반주기 수와 표본 반주기 수를 나타내며,  $s_{ijl}^2$ 은  $(i, j)$ 층에서  $l$ 곡목의 표본분산을 나타낸다. 즉,  $s_{ijl}^2 = 1/n_{ij-1} \sum_{k=1}^{n_{ij}} (y_{ijkl} - \bar{y}_{ijl})^2$ 이다.

### 3.5. 모수추정결과 분석

일반적으로 표본설계의 타당성 검증은 모집단의 특성인 모수의 참값을 모르기 때문에 불가능한 경우가 많지만 본 연구에서는 온라인 노래연습장에서 자동으로 로그데이터를 수집할 수 있어 가능하다.

구체적인 절차는 식 (3.2)을 이용하여 온라인 표본업소에서 수집한 로그데이터에서 추정된 히팅수와 전체 온라인 노래연습장의 로그데이터로부터 산출한 곡목별 히팅수 간의 상대오차를 비교하기위해서 1위-10위, 20위, 30위, 40위, 50위, 100위에 대한 계산결과를 제시하면 다음의 표 5와 같다.

표 5에서 보면 온라인 표본업소에서 수집한 로그데이터에서 추정된 히팅수와 전체 온라인 노래연습장의 로그데이터로부터 산출한 곡목별 히팅수 간의 상대오차가 상위 곡목에 대해서는 대부분의 경우에 허용오차(5%)보다는 작은 것으로 나타났고, 하위 곡목에 대해서도 대부분의 경우에 10% 이하로 나타났으므로 오프라인 노래연습장에서 곡목별 히팅수를 추정하기 위한 표본설계연구는 타당한 것으로 볼 수 있다.

## 4. 오프라인노래연습장의 모수추정

### 4.1. 자료수집

오프라인노래연습장의 모수추정은 로그데이터를 분석하여 가중치를 산출한 후에 2009년 5월 한



표 5: 온라인 노래연습장의 표본설계 타당성 검증(2009년2월데이터)

온라인모집단			온라인표본조사			상대오차
반주기수	순위	히팅수	히팅수	순위	반주기수	
22,806	1	384,364	358,961	1	110	0.066090997
17,478	2	290,469	284,017	2	83	0.022212353
15,413	3	219,796	211,775	3	74	0.036492930
17,058	4	215,490	202,145	4	81	0.061928628
16,441	5	213,306	194,962	5	77	0.085998519
14,622	6	173,745	148,150	7	71	0.147313592
14,628	7	170,151	147,506	8	66	0.133087669
14,041	8	156,663	150,049	6	67	0.042218009
16,034	9	153,504	135,847	9	79	0.115026319
15,154	10	142,282	131,281	10	75	0.077318283
13,540	20	94,060	89,381	18	68	0.049744844
11,470	30	70,818	66,588	32	49	0.059730577
19,972	40	62,268	59,713	39	93	0.041032312
11,765	50	52,528	49,193	58	52	0.063489948
9,913	100	36,629	31,433	135	44	0.141854814

달간(1일부터 31일까지를 기준)에 대한 1위에서 100위곡까지 곡명별 히팅수 지부별로 추정하는 것을 원칙으로 하지만 연구진행의 편의성을 위해서 전체적인 히팅수를 분석하였다.

오프라인 노래연습장의 자료수집 절차는 우선 지부별(14개)로 분류하여 업소 수(조사모집단)를 조사한 다음 표본배분은 지부별로 업소수를 기준으로 비례배분하고 표본업소는 계통추출법을 적용하여 선정한다. 표본으로 선정된 업소 내에서 제작사별로 반주기를 선정하여 로그데이터 수집기를 설치하고 로그데이터를 수집한다.

#### 4.2. 가중치 산출

과정\_1은 모집단의 반주기 수를 표본업소의 조사된 반주기 수로 나누어 계산한 것이고 최종가중치(과정\_2)는 일부 층에서 표본업소가 하나도 없으므로 벤치마킹보정가중치를 반영한 최종가중치다(Kish, 1965).

표본노래연습장에서 반주기 하나를 표본으로 선정하였으므로 표본추출과정은 노래연습장을 1차로 추출하고 선정된 표본업소에서 하나의 반주기를 2차로 추출하였으므로 2단계추출과정을 거쳐서 표본반주기를 추출한 것이다. 종합적으로 보면 각층별로 모집단의 반주기 수에서 표본업소 수(표본 반주기 수와 동일함)를 추출한 것으로 볼 수 있으므로 가중치(과정\_1)에서는 모집단 반주기 수를 표본업소 수로 나누어 계산한다.

표본업소가 하나도 없으면 가중치의 합계와 모집단 반주기 수가 일치하지 않을 수 있기 때문에 이를 보정하는 것이 벤치마킹가중치(과정\_2)이다. 벤치마킹가중치(과정\_2)는 가중치(과정\_1)에 표본업소수를 곱하여 합계를 계산한 후에 표본가중치합계의 반주기 수와 모집단의 반주기 수를 일치시키는 가중치를 벤치마킹보정가중치라 하고 이를 최종가중치로 한다.

위에서 언급한 내용을 종합하여 구체적인 절차로 설명하면 다음과 같다.

- (1) 업소별로 로그데이터 수집기를 설치한 일자가 상이하므로 로그데이터의 수집기간을 31일 기준으로 환산하는 가중치를 계산한다. 예를 들어, 수집기간이 9일이면 로그데이터를 9로 나누어 1일간 히팅수 산출한 후에 31일을 곱하여 31일간의 히팅수 계산할 수 있도록 가중치를 계산한다.

$$w_{1ij} = \frac{31}{d}, \tag{4.1}$$

표 6: 오프라인노래연습장의 가중치 산출결과

지부	제작사	모집단		표본		최종가중치 (과정 2)
		업소수	반주기수	업소수	반주기수	
강남	금영/태진	430/432	2,557/2,688	23/12	142/82	112.40/226.46
강북	금영/태진	1,102/1,968	6,414/11,356	50/54	305/352	129.69/212.61
강서	금영/태진	594/617	3,383/3,818	17/18	91/116	201.19/214.44
강원	금영/태진	323/373	1,822/2,144	7/13	31/83	263.15/166.74
경기	금영/태진	1,540/1,664	8,291/9,267	67/46	366/240	125.11/203.67
경남	금영/태진	2,224/203	11,460/1,058	51/2	258/14	227.18/534.82
경북	금영/태진	2,775/244	13,897/1,260	121/1	630/5	116.11/1273.86
부산	금영/태진	1,630/320	8,606/1,693	74/2	396/11	117.58/855.81
인천	금영/태진	340/1,823	1,906/10,704	6/56	41/292	321.16/193.25
전남	금영/태진	1,354/208	6,824/1,013	58/0	297/0	118.95/0
전북	금영/태진	694/66	4,180/412	21/0	124/0	201.24/0
제주	금영/태진	150/91	914/573	6/5	31/31	154.01/115.86
충남	금영/태진	1,449/446	7,492/2,454	59/13	327/74	128.38/190.85
충북	금영/태진	490/321	2,811/1,952	22/14	112/72	129.18/140.96
합계		23871	130949	818	4523	

여기서,  $d$ 는 로그데이터 수집기간이다.

- (2) 표본 추출률과 응답률을 반영하여 설계가중치를 계산한다.

$$w_{2ij} = \frac{M_{ij}}{m_{ij}}, \quad (4.2)$$

여기서,  $i$ 는 지부를 나타내고  $j$ 는 제작사를 나타내고  $M_{ij}$ 와  $m_{ij}$  각각  $i$ 지부내의  $j$ 제작사의 모집단 반주기수와 표본반주기 수를 나타낸다.(여기서는  $i$ 지부내의  $j$ 제작사의 표본업소가 최소한 1개 이상 조사된 것으로 가정하며 만일에 표본업소가 하나도 없는 층이 있을 경우에는 전체의 반주기수를 기준으로 가중치 합계와의 비를 벤치마킹보정가중치하여 이를 추출률과 응답률의 가중치에 곱하여 계산함)

- (3) 식 (4.2)로 주어진 가중치는  $i$ 지부내의  $j$ 제작사 층 내의 모든 표본업소는 동일한 가중치를 갖는 것으로 가정하였다.
- (4) 최종가중치는 식 (4.1)과 (4.2)를 곱하여 산출한다.

$$W_{2ij} = w_{1ij} \cdot w_{2ij} = \frac{31}{d} \cdot \frac{M_{ij}}{m_{ij}}. \quad (4.3)$$

2009년 5월(1일부터 31일까지를 기준)의 수집한 자료에 대해 지부별 제작사별로 가중치를 산출하면 다음 표 6과 같다.

#### 4.3. 모수추정

표본업소에 대한 가중치와 관찰값이 주어진 경우에는 전체 히팅수와 특정 곡목별 히팅수에 대한 추정식은 모 총계에 관한 추정식과 같은 형태를 갖게 될 것이다. 아래 식은  $i$ 지부의  $j$ 제작사 반주기에서 곡목명  $l$ 의 히팅수의 추정 값을 계산하기 위한 것이다 (Yates, 1981).

$$\hat{\tau}_{ijl} = \sum_{i=1}^{14} \sum_{j=1}^2 \sum_{k=1}^{n_{ij}} W_{2ij} \times_{ijkl}, \quad (4.4)$$

표 7: 온라인모집단의모수와 오프라인표본추정치의 비교(오프라인 기준)

오프라인표본추정치			온라인모집단모수		
히팅수	순위	반주기수	반주기수	순위	히팅수
923,739	1	2,936	13,180	63	63,333
850,125	2	2,294	37,202	1	405,727
699,706	3	1,961	17,944	28	99,350
672,914	4	2,048	12,482	98	50,938
670,420	5	2,033	13,740	71	60,756
278,870	50	818	9,676	268	23,300
187,525	100	678	6,607	541	12,254

표 8: 온라인모집단의모수와 오프라인표본추정치의 비교(온라인 기준)

온라인모집단모수			오프라인표본추정치		
반주기수	순위	히팅수	히팅수	순위	반주기수
37202	1	405,727	850,125	2	2,294
17670	2	390,377	632,489	6	1,569
16484	3	327,065	300,022	43	1,051
17511	4	275,866	453,477	15	1,427
17248	5	275,038	310,467	41	1,100
12721	50	71,404	79,588	262	244
12123	100	49,969	64,030	318	213

여기서,  $W_{2ij}$ 는 식 (4.3)에서 주어진 가중치를 나타내고,  $y_{ijkl}$ 는  $(i, j)$ 층 내의  $k$ 표본업소에서 수집한  $l$ 곡목의 히팅수를 나타낸다.

식 (4.4)에 주어진 추정량에 대한 분산 추정식은 일종의 층화비례추출법을 적용한 표본추출이므로 근사적으로 아래 식으로 계산할 수 있다.

$$\hat{V}(\hat{\tau}_{ijl}) = \sum_{i=1}^{14} \sum_{j=1}^2 w_{ij}^2 \left( \frac{1}{n_{ij}} - \frac{1}{N_{ij}} \right) s_{ijl}^2, \tag{4.5}$$

여기서,  $w_{ij} = N_{ij} / \sum_{i=1}^{14} \sum_{j=1}^2 N_{ij}$ 는  $(i, j)$ 층의 조사모집단에서 구성비를 나타내고  $N_{ij}$ 와  $n_{ij}$ 는  $(i, j)$ 층에서 각각 모집단 반주기 수와 표본 반주기 수를 나타내며,  $s_{ijl}^2$ 은  $(i, j)$ 층에서  $l$ 곡목의 표본분산을 나타낸다. 즉,  $s_{ijl}^2 = 1/(n_{ij} - 1) \sum_{k=1}^{n_{ij}} (y_{ijkl} - \bar{y}_{ijkl})^2$ 이다.

#### 4.4. 모수추정결과 분석

본래의 조사목적이 온라인과 오프라인의 특성을 비교분석하는 것이므로 노래연습장에서 추정된 오프라인의 1위-5위, 50위, 100위에 대한 곡목별 히팅수와 온라인에서 산출한 1위-5위, 50위, 100위에 대한 곡목별 히팅수를 비교하여 차이를 분석한다. 곡목별 히팅수의 비교는 표 7에서는 오프라인 기준으로 곡목별 추정된 히팅수와 온라인 모집단의 히팅수 비교하였으나 온라인의 곡목별 히팅수와 오프라인의 추정된 히팅수의 비교에서 차이가 현격하였다. 표 8에서는 온라인 히팅수를 기준으로 정렬한 후에 오프라인 표본업소에서 추정된 히팅수와 차이를 비교하였다.

#### 4.5. 온라인과 오프라인의 특성

표 7은 오프라인의 표본업소에서 수집한 로그데이터로부터 추정한 곡목별 히팅수를 근거로 순위를 제시하였는데 우선 상위 1위곡과 2위곡에 대한 히팅수를 살펴보면 1위곡(전통가요)의 히팅수가

표 9: 온라인모집단의모수와 오프라인표본추정치의 비교(온라인 기준)

방안	수록곡	온라인 히팅수	오프라인 히팅수
방안1	30%	70%	70%
방안2	20%	80%	80%

923,739회이지만 이곡에 대한 온라인 노래연습장에서 히팅수는 63,333회로 63위이고, 2위곡의 히팅수는 850,125회이지만 온라인노래연습장에서 산출한 히팅수는 405,727회로 1위를 차지하였으나 히팅수에서는 47.7%에 해당된다. 온라인노래 연습장과 오프라인 노래연습장에서 불리어지는 노래의 장르도 다를 뿐 아니라 히팅수에서도 현격한 차이를 보이고 있다.

표 8에는 온라인노래연습장에서 수집한 로그 데이터로부터 계산한 순위를 정리하였다. 1위곡은 오프라인에서는 2위이라는 것을 앞에서 설명하였고 2위곡의 히팅수가 390,377회이지만 이곡의 오프라인에서 추정한 히팅수 632,489회로서 6위에 해당되었으며 3위곡의 히팅수가 327,065회이지만 오프라인에 추정한 히팅수 300,022회로 43위에 해당되어 히팅수에서는 비슷하지만 순위에서는 현격하게 격차가 나는 것을 볼 수 있다.

다른 순위 곡에 대해서도 유사한 특성을 보여주고, 온라인노래연습장과 오프라인 노래연습장에서 불리어지는 노래의 장르와 히팅수에서도 현격한 차이가 있기 때문에 온라인노래연습장에서 수집한 로그데이터를 기준으로 계산한 곡목별 히팅수만을 기준으로 음악저작권료를 분배하는 현재의 규정은 개정을 위해서 검토되어야 할 것이다.

## 5. 분배방법의 개선안

### 5.1. 분배방안

현행 분배규정(2009년부터)에는 노래연습장은 수록곡에 20%, 온라인 로그데이터에 80%를 배분하고 있으나 본 연구에서 제안한 분배방안은 수록곡, 온라인 히팅수, 오프라인 히팅수 기준으로 일률적으로 배분하는 2개의 안을 제안하고자 한다.

방안1은 분배금액 총액을 우선 수록곡에 30%를 할당하고, 나머지 70%에 대해서는 온라인과 오프라인 각각의 히팅수 기준으로 배분하는 방안이고 방안2는 분배금액 총액을 우선 수록곡에 30%를 할당하고, 나머지 70%에 대해서는 온라인과 오프라인 각각의 히팅수 기준으로 배분하는 방안으로 정리하면 다음 표 9와 같다.

두 방안 모두 업소의 영업형태를 온라인과 오프라인으로 구분하여 곡목별 히팅수를 독립적으로 계산하고 이를 기준으로 분배금액을 계산함을 원칙으로 하였다.

### 5.2. 분배금액 계산절차

분배방안에 대한 곡목당 분배금액의 계산 절차는 다음과 같다 (Hansen 등, 1953).

- (1) 징수된 사용료를 지부별로 온라인과 오프라인으로 구분한 후에 제작사별 구성비를 기준으로 징수액의 30%(또는 20%)금액을 계산하여 이 금액을 기준으로 모든 수록곡에 동일하게 배분한다. 이를 수식으로 표현하면 아래와 같다.

$T_{ij}$ :  $i$ 지부의  $j$ 영업형태(온라인과 오프라인)업소의 징수금액  $P_{ijk}$ :  $i$ 지부의  $j$ 영업형태의  $k$ 제작사에 할당된 금액

여기서,  $\alpha_{ijk}$  는  $i$ 지부의  $j$ 영업형태의  $k$ 제작사 반주기 구성비이고,  $\beta$ 는 0.3 또는 0.2로 분배방안별로 수록곡에 대한 할당비율이다.

표 10: 업소형태별 방안별 징수금액의 분포(2009년 5월 기준)

구분	방안1				방안2			
	수록곡30%	온라인	오프라인	계	수록곡20%	온라인	오프라인	계
노래연습장	222,608,764	101,962,235	417,458,217	742,029,216	148,405,843	116,528,269	477,095,104	742,029,216

(2)  $i$ 지부의  $j$ 영업형태의  $k$ 제작사  $l$ 곡목명의 수록곡에 배분된 금액은 아래 식으로 계산 할 수 있다.

$$x^1_{ijkl} = P_{ijk}/X_{ijk}, \text{ 여기서 } X_{ijk} \text{는 } i \text{지부의 } j \text{영업형태 } k \text{제작사의 전체 수록곡 수이다.}$$

(3) 각 지부별로 온라인과 오프라인으로 업소를 구분하고 계산된 곡목별 히팅수 기준으로 분배금액을 아래와 같이 계산한다.

$H_{ij} = (1 - \beta) \cdot T_{ij}$  :  $i$ 지부의  $j$ 영업형태에서 히팅수 기준으로 분배할 금액은

$$x^2_{ijl} = H_{ij} \cdot \frac{Y_{ijl}}{\sum_{l=1}^L Y_{ijl}},$$

여기서,  $Y_{ijl}$ 은  $i$ 지부의  $j$ 영업형태에서  $l$ 곡목명의 히팅수이고,  $x^2_{ijl}$ 은  $i$ 지부의  $j$ 영업형태에서  $l$ 곡목에 분배된 금액이다.

(4)  $i$ 지부의  $j$ 영업형태에서  $l$ 곡목에 분배된 금액은 수록곡의 기본배분금액과 히팅수 기준으로 계산된 배분금액의 합계이며 아래 식으로 표현할 수 있다.

$$x^*_{ijl} = \sum_{k=1}^2 x^k_{ijkl},$$

여기서,  $k$ 는 제작사를 나타내는데 두 개의 제작사에 동시에 수록된 곡목은 합계하여 계산한다.

### 5.3. 개선방안 비교분석

노래연습장의 분배금액을 기준으로 상위 100위곡까지 분배금액을 비교하는데 사용한 데이터는 2009년 05월의 표본업소에 대한 조사 자료와 협회에서 징수한 징수금액을 기준으로 수록곡 할당액, 온라인, 오프라인의 히팅수에 따른 분배금액으로 구분하여 각각 산출한 후 합계를 이용하여 방안1과 방안2에서 비교하고자 한다.

계산과정에서 필요한 자료를 살펴보면 2009년 05월의 협회에서 노래연습장에 대한 징수금액은 742,029,216원이고, 업소형태별로 반주기수 분포현황을 보면 온라인의 반주기수는 32,921개(19.63%), 오프라인 반주기수는 134,756개(80.37%)이며 반주기 제작사별 수록곡수는 태진 18,577곡, 금영 18,001곡이다.

징수총액을 방안별로 분류한 후 계산절차에 의거하여 산출하면 다음 표 10과 같다.

표 10에서보면 방안1의 수록곡 할당액은 222,608,764원으로 방안2보다 수록곡에 총액의 10%를 많이 배정하기 때문에 약 7천만원이 증액된다. 이를 구체적으로 보기위해 방안별로 수록곡당 단가를 산출하면 방안1은 12,172원이고 방안2는 8,114원으로 방안1의 수록곡당 할당액은 방안2보다 4,058원이 증액된다. 수록곡에 대해서 방안1(30%)과 방안2(20%)별로 배분한 후에 노래연습장 히팅수의 단가를 산출하면 다음 표 11과 같다.

표 11에서보면 방안별 히팅수 단가의 차이는 미미하나 히팅수의 단위가 크기 때문에 히팅수 산출시 정확성을 기해야한다.

표 11: 히팅수당 단가 산출 결과

구분	온라인			오프라인		
	히팅수	방안1	방안2	히팅수	방안1	방안2
노래연습장	33,860,558	3.0112391	3.4414161	103,006,060	4.0527540	4.6317188

표 12: 방안1과 방안2의 배분금액의 비교

구분	순위	차이: 방안2 - 방안1				합계비: 방안2/방안1
		수록곡	온라인	오프라인	합계	
노래연습장	1	-4,057	174,534	492,192	662,669	1.1416
	2	-4,057	27,245	534,813	558,001	1.1414
	3	-4,057	167,931	366,189	530,063	1.1413
	4	-4,057	42,738	405,105	443,786	1.1410
	5	-4,057	26,136	388,149	410,228	1.1409
	50	-4,057	10,035	185,965	191,943	1.1387
	100	-4,057	38,772	85,175	119,890	1.1363

표 13: 노래연습장의 현행과 제안 분배방안과의 비교

순위	현행분배액		제안방안 분배액		분배액 차이			
	분배1 (수록곡30%)	분배2 (수록곡20%)	방안1 (수록곡30%)	방안2 (수록곡20%)	분배1 - 방안1	분배1 - 방안2	분배2 - 방안1	분배2 - 방안2
1	6,177,396	7,053,839	4,679,259	5,341,928	1,498,137	835,468	2,374,580	1,711,911
2	1,084,590	1,233,488	3,946,567	4,504,568	-2,861,977	-3,419,978	-2,713,079	-3,271,080
3	5,988,261	6,837,677	3,751,011	4,281,074	2,237,250	1,707,187	3,086,666	2,556,603
4	1,589,106	1,797,976	3,147,073	3,590,859	-1,557,967	-2,001,753	-1,349,097	-1,792,883
5	1,005,592	1,143,196	2,912,168	3,322,396	-1,906,576	-2,316,804	-1,768,972	-2,179,200
50	416,732	461,149	1,384,173	1,576,116	-967,441	-1,159,384	-923,024	-1,114,967
100	1,355,038	1,542,574	879,799	999,689	475,239	355,349	662,775	542,885

분배금액 계산절차에 의거 노래연습장의 곡목별 히팅수를 기준으로 방안1과 방안2에 대해서 분배금액을 계산하였는데 방안1과 방안2의 배분금액의 비교를 위해서 1위-5위, 50위, 100위 곡에 대해서 각 항목 및 합계의 차이와 합계비를 산출해보면 다음의 표 12와 같다.

표 12에서 방안1에 대해서 방안2를 비교해 보면, 1위곡은 수록곡 할당액에서 4,057원이 감소하나 온라인에서 174,534원이 증가하고 오프라인에서 492,192원이 증가하여 합계는 662,692원이 증가하면서 14%가 상승한다.

또한, 50위곡은 수록곡 할당액은 4,057원이 감소하나 온라인에서 10,035원이 증가하고 오프라인에서 185,965원이 증가하여 합계는 191,943원이 증가하면서 13%가 상승한다.

물론 100위곡도 수록곡 할당액은 4,057원이 감소하나 온라인에서 38,772원이 증가하고 오프라인에서 85,175원이 증가하여 합계는 119,890원이 증가하면서 13%가 상승한다.

마지막으로 현행분배금액과 본 연구에서 제안한 방안1 및 방안2의 배분금액의 비교를 위해서 노래연습장의 상위 1위, 5위, 50위, 100위곡에 대한 합계를 산출하여 정리한 결과를 노래연습장은 다음 표 13에 정리하였다.

표 13에서 보면 현행분배규정과 방안1 및 방안2를 적용한 곡목별 배분금액을 산출한 결과분석을 보면 분배금액을 기준으로 상위 1위는 현행분배규정에서는 6,177,396원(7,053,839원), 방안1(수록곡 30%)에서는 4,679,259원, 방안2(수록곡 20%)에서는 5,341,928원으로 1,498,137원이 제안한 분배방안에서 감소하는 것으로 나타났으나 2위(전통가요)는 현행분배규정에서는 1,084,590(1,233,488), 방안1(수록곡 30%)에서는 3,946,567원, 방안2(수록곡 20%)에서는 4,504,568원으로 새로운 분배방안에

서는 분배금액이 2,861,977원이 증가하는 것으로 나타나서 전통가요분야에서는 새로운 분배방안에서 배분금액이 대단히 증가됨을 알 수 있다.

#### 5.4. 최적개선안 선정

노래연습장에서 히팅수가 높은 곡목일수록 히팅수에 대한 배분비율이 높은 방안2에서 분배금액이 증가하는 현상을 보이고 히팅수가 작은 곡목은 방안1에 의한 분배금액이 크게 나타난다. 또한 1위곡과 100위곡을 비교해보면 방안1에 비해 방안2의 배분액의 상승률은 13%~14%로 비슷하나 상승액의 차이는 무려 6배 이상의 차이를 보여서 히팅수가 많은 곡은 분배금액이 증가하고 히팅수가 작은 곡은 분배금액이 감소하는 富益富 貧益貧의 현상을 더욱 가중 시킨다.

종합해보면, 방안1과 방안2의 차이는 수록곡의 배분비율을 30% 또는 20%로 하는 것으로 분배금액에서 현격한 차이가 없으나 저작권사용 빈도수인 곡목별 히팅수를 얼마나 반영할 것인가라는 측면에서는 중요한 의미가 있다. 그러나 규정개정의 충격이나 혼란을 최소화 할 수 있고 곡목별로 장르에 무관하게 현행 분배규정에 의한 분배금액과 제안한 분배방안에 의한 분배금액의 차이가 상대적으로 적은 방안1(수록곡 30%반영)을 본 연구 결과로 제안한다.

### 6. 결론

음악저작권료의 징수와 분배의 실태를 파악하기 위해서 현재 적용되고 있는 음악저작물 사용료 분배규정을 검토하였으며 또한 저작권사용료의 징수실태를 파악하기 위해서 모집단을 분석하였다. 모집단의 분석은 모집단을 구성하는 노래연습장의 명부를 사용하였고, 모집단의 층화변수를 도출하기 위해서는 2008년 12월에 온라인업소에서 수집된 로그데이터의 히팅수와 기기 당 평균 히팅수에 대해서 다중비교와 분산분석을 하여 층화변수로는 시도, 업소형태, 시도규모 및 업소규모 등을 선정하였다.

표본크기의 결정도 2008년 12월의 로그데이터에 층화비례추출법을 적용하고 95%의 신뢰도와 5%의 신뢰수준으로 다양한 허용오차크기에 대한 시뮬레이션을 통해서 검토하고 전문가, 한국음악저작권협회 관계자와 담당공무원의 협의를 거쳐서 노래연습장 1,250개 업소로 확정하였다.

표본배분은 앞서 도출된 층화변수를 순서대로 적용하여 비례배분법으로 할당하였으며 모집단의 규모가 30이하인 층은 통합하여 모든 층에서 표본이 추출될 수 있도록 하였다. 각 층별로 표본을 할당한 후에는 표본의 대표성을 제고하기 위해서 업소의 주소코드를 기준으로 정렬한 후에 계통추출법으로 표본업소를 선정하였다.

현재 분배규정은 온라인업소에서 수집한 로그데이터를 근거로 곡목별로 히팅수를 계산하고 이를 기준으로 징수액의 70%를 분배하고 있다. 이는 온라인업소에서 불리어지는 곡목들이 오프라인에서 불리어지는 곡목들의 특성과 상이함을 무시하고 있기 때문에 표본설계연구를 통해서 오프라인 업소에서 불리어지는 곡목들의 특성과 히팅수를 추정하기 위해서 표본업소를 계통추출법으로 선정한 후 로그데이터 수집기를 반주기에 부착하여 로그데이터를 수집, 분석하여 곡목별 히팅수를 추정하 후 온라인 결과와 비교하고 이를 근거로 분배방안을 제안하며 제안된 방안별로 분배금액을 산정하여 방안에 대한 타당성을 검토하였다.

표본설계에 의해서 선정된 표본업소가 조사대상업소의 특성을 대표할 수 있는지를 분석한다면 표본업소에서 수집된 로그데이터로부터 히팅수를 계산한 결과를 근거로 한 분배방안은 정당성이나 타당성이 확고해질 수 있으므로 온라인업소데이터를 이용하여 분석하였다. 표본설계연구당시에 온라인 업소도 오프라인 업소와 동일한 알고리즘을 이용하여 추출하였기 때문에 2009년2월에 수집한 노래연습장의 온라인 업소들의 로그데이터를 이용하여 표본업소로 선정된 133개 업소로부터 추정된 곡목별 히팅수와 전체 업소 데이터로부터 계산한 곡목별 히팅수를 비교한 결과에서 상위 20위곡까지 중에서

상대오차비율이 10%미만이 70%정도인 특성을 보이므로 표본설계연구에서 표본업소를 선정하는 절차에 대한 정당성은 있다고 할 수 있을 것이다.

노래연습장은 오프라인과 혼합형으로 구분하여 로그데이터를 분석하였는데 특성이 오프라인과 혼합형은 유사하고 혼합형의 표본수가 33개로 작기 때문에 곡목별 히팅수 산출에서는 분석하였으나 분배액의 방안비교에서는 혼합형은 온라인과 오프라인으로 구분하여 통합적으로 분석하였다. 노래연습장의 오프라인 업소에서 추정된 상위 100위곡까지 곡목별 히팅수의 분석에서 오프라인 업소에서는 전통가요중심으로 불리어지고 있으며 반면에 온라인에서는 최신 곡 중심으로 젊은 층에서 좋아하는 장르 중심의 노래들이 불리어지고 있음을 알 수 있었다. 또한 히팅수에서도 오프라인 1위곡이 온라인에서 1위곡의 2배 이상의 차이를 보이고 있으므로 현행의 온라인 로그데이터만을 이용한 분배방법은 개정할 필요가 있음을 시사하고 있다.

분배방안의 제안에서 분배원칙은 수입이 발생하는 요소별로 분배비율을 결정하는 것이 모든 관계권리자를 설득할 수 있는 적절한 방안으로 생각하여 온라인 업소와 오프라인 업소를 구분하여 징수된 금액을 분배하는 방안을 제안 하였다. 방안1에서는 수록곡에 30%를 우선 할당하고 나머지 70%는 업소형태별로 추정된 히팅수를 근거로 배분하는 방안이고 방안2는 수록곡에 배분하는 비율을 20%로 낮추고 나머지 80%는 방안1과 동일한 형식을 적용한 방안이다. 방안1 및 방안2와 현행분배규정을 적용한 곡목별 배분금액을 산출한 결과분석에서 분배금액을 기준으로 상위 1위곡은 150만 원 정도가 제안한 분배방안에서 감소하는 것으로 나타났으나, 2위(전통가요)는 제안한 분배방안에서 280만 원 정도 증가하는 것으로 나타나서 전통가요분야에서는 새로운 분배방안에서 배분금액이 획기적으로 증가됨을 볼 수 있다.

방안1과 방안2의 차이는 수록곡의 배분비율을 30% 또는 20%로 하는 것으로 분배금액에서 현격한 차이가 없을 수 있으나 저작권사용 빈도수인 곡목별 히팅수를 얼마나 반영할 것인가라는 측면에서는 중요한 의미가 있으며 또한 곡목별로 장르에 무관하게 현행 분배규정에 의한 분배금액과 제안한 분배방안에 의한 분배금액의 차이가 상대적으로 적은 방안1을 최적방안으로 제안한다.

앞으로 업소들의 영업형태가 온라인인지 오프라인지가 중요한 관심대상이므로 관리차원에서 모집단관리를 정확하게 할 수 있는 시스템을 갖추는 것이 한국음악저작권협회의 중요한 업무가 될 것이다. 또한 표본업소에서 로그데이터를 수집하는 것도 곡목별로 히팅수를 추정하는데 핵심사항이므로 오류를 최소화하도록 표본업소 관리지침을 작성하여 통계품질관리에 유의해야 할 것이다 (통계청, 2008).

## 참고 문헌

- 박홍래 (2000). <통계조사론>, 영지문화사.  
 이계오, 박진우, 이기재 (2007). <표본조사론>, 한국방송통신대학교 출판부.  
 통계청 (2008). <통계품질관리 이렇게 합니다>, 통계품질진단 핸드북.  
 Biemer, P. P. and Lyberg, L. E. (2003). *Introduction to Survey Quality*, John Wiley & Sons, New York.  
 Cochran, W. G. (1977). *Sampling Technique*, 3rd edition, John Wiley & Sons, New York.  
 Hansen, M. H., Hurwitz, W. N. and Madow, W. G. (1953). *Sample Survey Methods and Theory*, Wiley, New York.  
 Kish, L. (1965). *Survey Sampling*, Wiley, New York.  
 Yates, F. (1981). *Sampling Methods for Censuses and Surveys*, 4th edition.



# Study of Sample Design for Improving the Method of Distribution of Singing Rooms Royalty

Kay O Lee<sup>1,a</sup>, Jungbai Choi<sup>b</sup>

<sup>a</sup>Department of Business statistics, Hannam University

<sup>b</sup>Department of Basic science, Airforce Academy

---

## Abstract

Sample design is studied to estimate the number of records from singing rooms. The definition of survey population, stratification, sample size and allocation of sample are discussed with using the collected real data. The validity of sample design is testified by the log data collected from online singing room which is sampled with suggested design. Estimation of the number of hitting of off-line singing room is proposed and calculated from the log data collected from sampled off-line singing rooms. The difference of online and off-line singing rooms is analyzed and the method of distribution of singing room royalty is suggested.

**Keywords:** Definition of survey population and stratification, online and off-line singing room, estimation of the number of hitting.

---

---

<sup>1</sup> Corresponding author: Professor, Department of Business statistics, Hannam University, 133 Ohjung-Dong, Deaduk-Gu, Daejun 306-791, Korea. E-mail: kayolee@hnu.kr